JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 19, No. 8 AUGUST, 1948

In This Issue

		Page
Theory of the Radio Detection of Meteors	LAURENCE A. MANNING	689
Aids to Analysis of Patterns Obtained in the Diffraction of Electrons by Gases	H. J. YEARIAN AND W. M. BARSS	700
Some Observations on the Formvar Replica Method, with a Note on Glass Surfaces	BEATRICE M. DEACON, S. G. ELLIS, W. G. CROSS, AND R. S. SENNETT	704
Pseudostructures in Electron Microscope Specimens	JOHN H. L. WATSON	713
Range-Finder Testing Laboratory	IRVINE C. GARDNER	721
Testing of Range Finders	IRVINE C. GARDNER	729
Method for the Evaporation of Alloys	LOUIS HARRIS AND BENJAMIN M. SIEGEL	739
Statistical Factors Affecting the Intensity of X-Rays Diffracted by Crystalline Powders	LEROY ALEXANDER, HAROLD P. KLUG, AND ELIZABETH KUMMER	742
Methods for Measuring Dynamic Mechanical Properties of Rubber-Like Materials	A. W. NOLLE	753
Nucleation	J. C. FISHER, J. H. HOLLOMON, AND D. TURNBULL	775
Analytical Discussion of the Construction of Pole Figures	JOHN K. WOOD	784
Electron Microscope Study of Gold Smoke Deposits	LOUIS HARRIS, DAVID JEFFRIES, AND BENJAMIN M. SIEGEL	791

Letters to the Editor:

Prebunching: A Method Controlling Synchrotron Radial Oscillation Amplitudes	D. C. DEPACKH AND M. BIRNBAUM	795
Limiting Sensitivity of a Microwave Spectrometer	C. H. TOWNES AND S. GESCHWIND	795
Gyrograms	HENRY O. FUCHS	796
Note on Reflection from Dielectric Structures in Wave Guide	J. SHMOYS	797
Theoretical Criterion for Streamer Advance in an Electrical Field	LEONARD B. LOEB AND ROBERT A. WIJSMAN	797
Notes on Electron Diffraction with the Universal Electron Microscope	R. E. DUNAWAY AND H. M. BARTON	799

In Every Issue

Calendar of Meetings	728
New Books	800
New Booklets	801
Here and There	801
Index to Advertisers	xxx

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND
JOHN VON NEUMANN

Publication of the

American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

CLEVELAND NORCROSS, *Assistant Director*

RUTH F. BRYANS, *Publication Manager*

THEODORE VORBURGER, *Business Manager*

JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

GEORGE R. HARRISON, *Chairman*

WALLACE WATERFALL, *Secretary*

GEORGE B. PEGRAM, *Treasurer*

LUIS W. ALVAREZ
ROBERT B. BRODE
E. U. CONDON
A. J. DEMPSTER
W. F. FAIR, JR.
HARVEY FLETCHER
R. C. GIBBS
W. V. HOUSTON
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
A. H. PFUND
JOHN T. TATE
M. W. WHITE
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members and associate members of the American Institute of Physics	$6.00	$7.00
To all others	8.00	9.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 18, 1947—$132.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1948, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 19, Number 8 August, 1948

The Theory of the Radio Detection of Meteors

LAURENCE A. MANNING
Stanford University, California

(Received August 20, 1947)

The methods of observation are described both for "bursts" of signal strength and "whistles" caused by reflection of radio waves from meteoric ionization. Consideration is given to the geometry of reflection, and it is shown that meteoric velocity and range can be accurately determined by combined pulse and whistle observations. The possibility of determining the direction and position of a meteor path by observations at three receiving locations is examined, and a method for interpreting such data is developed. An investigation is made of the fraction of meteors capable of producing bursts, and the important effect of the meteoric radiant is emphasized.

I. INTRODUCTION

As meteors pass through the upper atmosphere they leave behind them a trail of ionized air of surprising dimensions. Skellett[1] has been notable among those pointing out that a typical meteor, traveling at a velocity of 40 km/sec. and having a mass of a small fraction of a gram, possesses sufficient energy to produce a column of ionization perhaps one kilometer in diameter and one hundred kilometers long. He has shown, moreover, that the ionization in this column can be great enough to be effective in reflecting radio signals.

There are two types of reflections which may be obtained from these ionization trails. The first type is generally known as a "burst" of signal strength, and Pierce[2] has suggested that bursts result when the transmitted signal is reflected from the cylinder of meteoric ionization in such a way that the incident and reflected rays make supplementary angles with the meteor's path. The signals seen on radar oscilloscopes are of this type.[3,4] The second type of reflection[5] is of a transient nature, and apparently results when a continuous wave signal impinges upon the head of the rapidly advancing column of ionization. Because of the motion of the reflecting point, the returned signal suffers a Doppler shift in frequency and may be combined with a portion of the original transmitted energy to produce a beat frequency which ordinarily lies within the audible range. Reflections of this second type produce what are termed "meteor whistles."

When a meteor is observed to be accompanied by both a whistle and a burst, the burst is in-

[1] A. M. Skellett, "The ionizing effect of meteors in relation to radio propagation," Proc. I.R.E. **23**, 132–149 (1935).

[2] J. A. Pierce, "Abnormal ionization in the *E*-region of the ionosphere," Proc. I.R.E. **26**, 892–902 (1938).

[3] R. Bateman, A. G. McNish, and V. C. Pineo, "Radar observations during meteor showers—9 October 1946," Science **104**, 434–435 (1946).

[4] J. S. Hey and G. S. Stewart, "Derivation of meteor stream radiants by radio reflexion methods," Nature **158**, 481–482 (1946).

[5] Chamanlal and K. Venkatamaran, "Whistling meteors—a Doppler effect produced by meteors entering the ionosphere," Electrotechnics **14**, 28–40 (1941).

variably the much stronger reflection. The whistle signal is weaker principally because the head of the ionization column, where the whistle reflection occurs, is newly formed and has not had time to expand to its full size. An additional weakening of the whistle signal results when the incident ray makes a small angle with the meteoric path, since then the equivalent reflecting surface of the ionization cloud exhibits curvature not only in a circumferential direction, but also has roundness of the nose.

II. THE CONDITIONS OF OBSERVATION

Reflection from meteor trails may be most easily observed if a powerful short wave broadcast station is located about one hundred miles from the receiving position.[6] It is, of course, essential that the non-meteor wave at the receiver be made very weak, so that the extremely small signals reflected from the meteors will not be overwhelmed. If the transmitter is located beyond the range of the ground wave, the principal component of signal strength remaining is a result of scattering. In a comprehensive study of the nature of these scattered signals,[7] Eckersley has shown that they originate in the *E*-layer at a distance from the transmitter equal to or slightly greater than the skip distance for the *F*-layer. Their magnitude will hence decrease as the skip distance increases, and when, because of either the use of high enough frequency or the presence of a low enough *F*-layer ion density, propagation is no longer possible to *E*-layer heights via the *F*-layer, the scattered signals become extremely small.

Because only a very weak background of scattered signal can be tolerated when observing meteors, it is important to choose a broadcast station on a frequency so high that the skip zone is far removed. If the frequency used is greater than three and one-half times the critical frequency of the *F*-layer, the skip zone vanishes into the "ionospheric horizon," which is the optical horizon for an observer situated at twice the layer height.

If, now, a transmitter satisfying the above requirements is monitored, one should be able to observe both meteor whistles and signal strength bursts due to meteors. The whistles generally commence at a frequency in the neighborhood of one-thousand cycles and decrease in pitch for as long as a second. In some instances the whistle dies out before reaching a very low pitch, but in those cases in which the pitch drops to substantially zero, a burst invariably follows the whistle. The bursts are much stronger than the

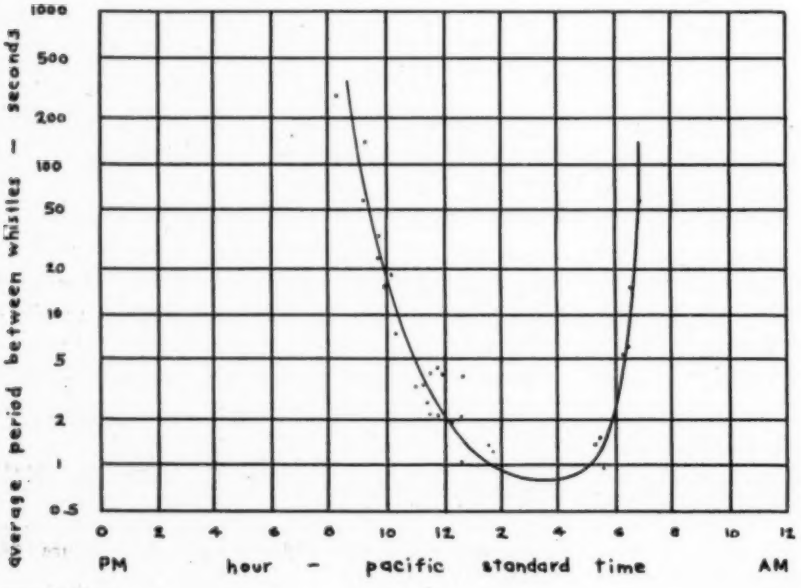


FIG. 1. Observed frequency of occurrence of whistles on a 19-meter broadcast station September 16–17, 1946.

[6] O. G. Villard, "Listening in on the stars," Q.S.T. **30**, 59 (1946).
[7] T. L. Eckersley, "Analysis of the effect of scattering in radio transmission," J. Inst. Elec. Eng. **86**, 548–567 (1940).

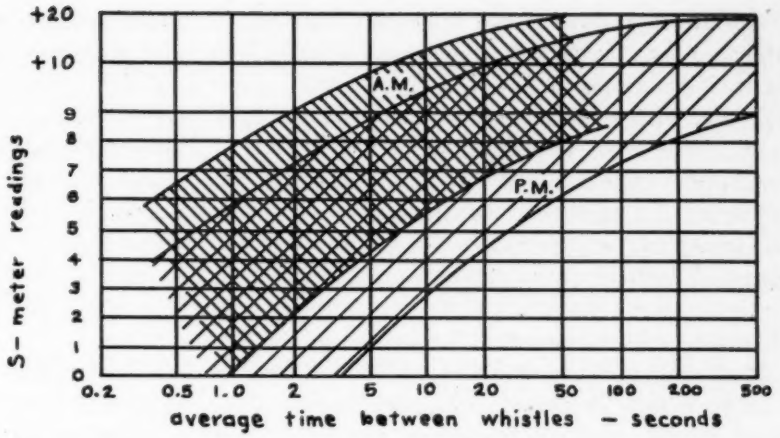


FIG. 2. The relation between long scatter signal strength and the observed rate of whistle detection.

whistles when they occur and may last several seconds, or only a small portion of a second. Occasionally a whistle of rising pitch will occur independently or, sometimes, following a burst. On the night of September 16, 1946, at 11:20 P.M., observation of a 15.34-mc broadcast station of about 50-kw power indicated whistles of falling pitch to predominate in the ratio of 3.3 to one, while the total rate of detection was about thirty per minute.

The number of whistles that can be heard depends very strongly upon the level of the scattered and direct signal with which the meteor reflections compete. Figure 1 gives some typical results from observations of a short-wave broadcast transmitter of about 100-kw power on a wave-length of 19 meters. In this band a strong scattered signal was obtained during the early evening hours, so that the average time between observed whistles cannot be taken as any indication of the actual diurnal variation of meteor frequency. In Fig. 2 is plotted the observed range in which the signal strength meter fluctuated due to the overriding scattered signal, as a function of the average time interval between whistles. Few meteors were detected until the long scatter signal became weak. It will be seen that in the early morning, when the signal strength began to rise, the average periods between whistles were less than for the corresponding signal strengths in the evening. This difference in observed whistle frequency under these conditions of equivalent interference can probably be taken as an indication that meteors were actually considerably more numerous in the early morning. In way of explanation of the signal strength readings plotted in Fig. 2 is Fig. 3. The latter figure shows the relative frequency with which the signal strength meter attained various readings as relative maxima under one particular condition of reception. For use in plotting Fig. 2 this signal variation was evaluated as S (0–8); the other signal strength conditions were appraised similarly.

It is interesting to estimate the magnitude of the meteors which are represented by Fig. 1. If we assume the observed meteors to impinge on the atmosphere over as large an area as 100,000 square kilometers and if we accept Watson's values for the rate of meteoric incidence,[8] we must conclude that this elementary technique is capable of indicating meteors of about the eighth visual magnitude. The unaided eye is similarly effective up to about the fourth magnitude. The enormous relative sensitivity of radio methods is therefore apparent.

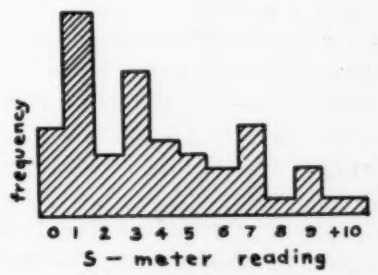


FIG. 3. Typical distribution of relative maxima of signal strength due to long scatter.

[8] F. G. Watson, *Between the Planets* (The Blakiston Company, Philadelphia, 1941).

Monitoring short wave broadcast stations is the simplest way to detect meteors by radio. For scientific purposes it is evidently not very satisfactory. There are several serious objections, of which the most important is the difficulty of interpreting the count of whistle frequency in the presence of a markedly varying scattered signal. If broadcast stations were available on frequencies sufficiently high so that *F*-layer propagation to the *E*-layer would be impossible at any distance, this foremost difficulty would be resolved. But it is not to be expected that broadcast stations will appear on frequencies such that long distance communication is completely impossible, unless we consider very high frequencies such as are used in frequency modulation work. It will be shown in subsequent papers that such very high frequencies are quite undesirable for whistle observations. Broadcasting stations have the additional disadvantage of possessing program modulation. Since most whistles produce a weak beat, they are blanketed to a considerable extent by the music or other program material. Broadcast stations have operating schedules which often cause them to sign off the air during the hours of greatest interest. In addition, unless one happens to be situated in a favorable geographical location, suitable transmissions from within the skip zone may not be available.

In order to dispose of the difficulties encountered when using commercial stations it is desirable to generate a continuous wave signal especially for this purpose.[9] The frequency used should ideally be the lowest for which little or no scattered signal is received, and the power should be as high as can be procured. The geometry involved in interpreting the characteristics of the reflected signal is greatly simplified if the transmitter and receiver are located close together. Under these conditions there may be a very strong ground wave which must be nulled out by the use of such a device as a universally mounted dipole antenna.

If only bursts are to be studied, it is possible to separate the meteor signals from the ground and scattered waves by means of a pulse technique. Ranges may be determined by this method, and transmitted frequencies in any desired portion of the spectrum may be used. Examples of the detection of meteors by the pulse technique are found, first, in ionospheric virtual height *versus* time records,[10] and second, in radar observations.[3] Pulse and continuous wave experiments both supply information not easily obtainable in any other manner and so complement each other.

III. GEOMETRY OF WHISTLE AND BURST PRODUCTION

A meteor whistle is the beat frequency resultant of a wave of transmitted frequency and of a reflected signal with its Doppler frequency shift. A Doppler shift in the received signal frequency will occur whenever the total optical path length from transmitter to reflecting point is changing. If the path length is constant as the meteor traverses its course, there will be no shift. The locus of all possible meteor positions giving a constant total path length is an ellipsoid of revolution having the transmitter and receiver as focii. When the reflection point enters these ellipsoids the reflected wave is shifted to a higher frequency, but the extent of the shift decreases with time for a constant linear velocity of the reflecting point; down whistles are thus produced. Up whistles are associated with velocities out of the ellipsoids. If meteor whistles are counted with the aid of a selective receiver tuned slightly to one side of the transmitted carrier frequency, it has been found that with the receiver tuned high, down whistles are heard predominantly, while with the receiver tuned low, mainly up whistles are heard.

The sequence of events leading to the production of a reflection can be understood with the aid of Fig. 4. Assume for sake of explanation that the characteristics of the atmosphere are identical along the whole path of the meteor, so that behind the meteoric particle there is an ionized region which effectively travels with meteoric velocity and unchanging shape. Near the meteor the column has not had time in which to expand to a large size, but at points separated appreciably in time and path length from the nose diffusion causes the column to grow in radius. Eventually, of course, the ionization density

[9] L. A. Manning, R. A. Helliwell, O. G. Villard, Jr., and W. E. Evans, Jr., "On the detection of meteors by radio," Phys. Rev. **70**, 767–768 (1946).

[10] J. A. Pierce, "A note on ionization by meteors," Phys. Rev. **59**, 625–626 (1941).

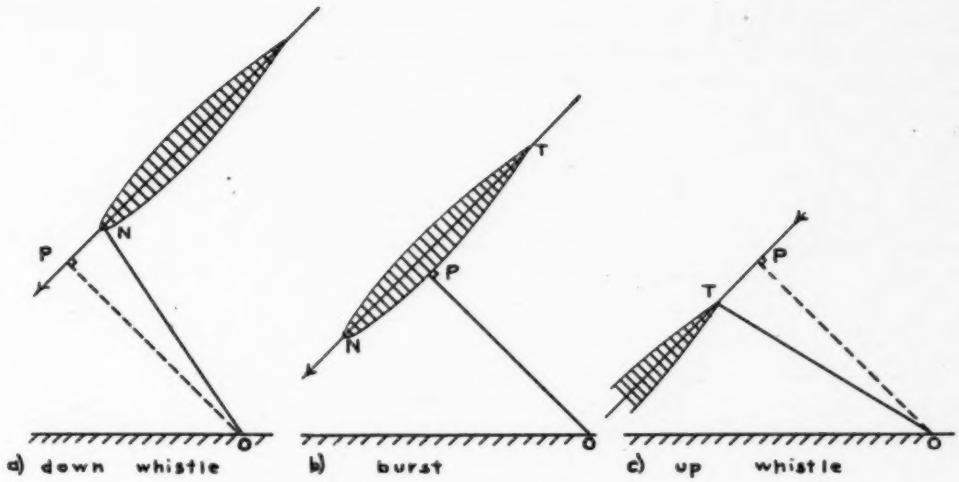


FIG. 4. The geometry of whistle production.

diminishes because of diffusion and recombination of the charge until the effective reflection radius approaches zero. The maximum radius of the cloud is greatly exaggerated in the sketches of Fig. 4, since typically the diameter is less than a percent of the length.

If a transmitting and receiving installation is located at an observatory 0, as in Fig. 4, the wave path which corresponds to the reflection condition of geometrical optics is the one (or ones) which meets the effective reflection surface of the cloud at right angles. The point of reflection is also the point on the reflection surface which is nearest to the observatory. In Fig. 4a is shown the ionization column before the meteor has arrived at the point P corresponding to a normal from the meteoric path to the observatory. Since the radius of the ionization column is exceedingly small in comparison with path lengths, the reflection point of Fig. 4a is located within a few meters of the meteor. The rate of change in path length to the observatory is a very exact measure of the meteoric velocity under the condition in Fig. 4a. Since the meteor is approaching the observatory a down whistle is produced.

When the distance $\langle PN \rangle$ is large, the reflection point is near the very tip of the ionization column, and this tip is so sharp that only very weak reflection can be obtained. As the meteor comes towards the perpendicular point P, the whistle signal gets stronger and stronger; often the radial velocity of the meteor with respect to the observatory has been reduced to about 10 or 20 percent of the meteor's linear velocity before the reflected signal is strong enough to be detected. The whistle declines in pitch almost entirely due to geometrical reasons under the conditions of Fig. 4a; the deceleration of the particle is not more than a few percent in the entire life of the meteors, and is certainly minor over the portion of the path which serves to produce strong whistles.

Eventually the nose of the ionization column passes the point of perpendicularity P of Fig. 4a, so that the next figure becomes applicable. Since the point on the reflecting surface nearest the observatory is now well away from the nose of the column, reflection is obtained from a cloud which has had time to diffuse to relatively large radii. As a result, the reflected signal strength starts at the low value corresponding to the whistles of Fig. 4a, and rises until the reflection point is at the portion of the cylinder having a maximum radius. There is no longer the direct connection between the velocity of the meteor and the observatory path length which determined the Doppler pitch of the whistle, since reflection is no longer confined to the nose of the column. The ionization column is exceedingly thin, so that the reflected ray must be very nearly perpendicular to the meteoric path. The only Doppler shift experienced by this strong burst reflection is caused by the radial expansion of the trail associated with diffusion. There is, of course, a gradual transition between the burst and whistle conditions, and at this time the Doppler

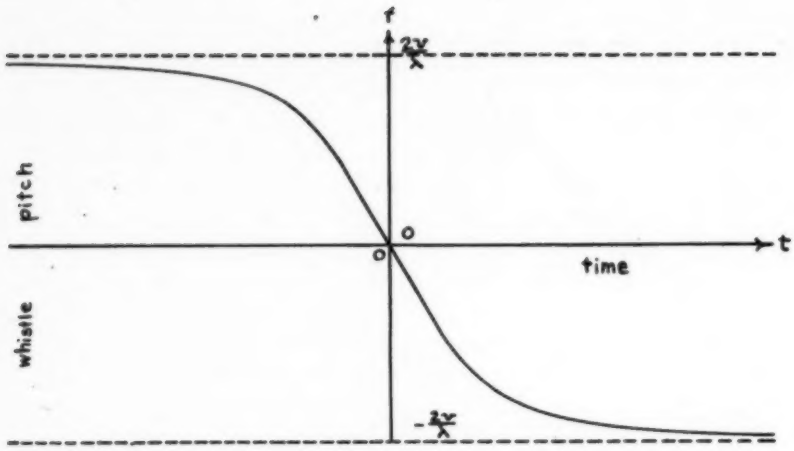


FIG. 5. Whistle pitch *versus* time for reflection from a sphere traveling in a straight line with a constant velocity.

pitch is determined by both the radial velocity of the meteor and the rate of trail expansion. Experimentally it is found that the whistle pitch does not begin to show alteration due to diffusion expansion until the Doppler on a 30-mc carrier has been reduced to perhaps 100 c.p.s. In a future paper the writer will show how analysis of the actual whistle pitch curve in the transition region permits calculation of the exact shape of the nose of the ionization column.

If finally the tail T of the column passes by the perpendicular point P, the production of an up whistle is possible. They are of relatively infrequent occurrence for geometrical reasons to be discussed in a later paper.

IV. DETERMINATION OF VELOCITY AND RANGE FROM WHISTLE ANALYSIS

Since the pitch of down whistles is determined by the radial velocity of the nose of the ionization column except for a fraction of a second preceding the burst, we shall investigate the theoretical whistle pitch curve corresponding to the geometry of Fig. 4a. Let the perpendicular distance $\langle OP \rangle$ from the observatory to the meteor path be called D. Then D is the range which will be measured when observing bursts with the pulse technique, as is obvious from Fig. 4b. Assume the meteor to progress towards P with a velocity v. We shall take v to be constant; this approximation is justified on the basis of the small photographically determined decelerations experienced by meteors. Now we wish to find the Doppler shift produced by the meteor's velocity when the meteor is at N. This shift will be identically equal to the whistle frequency f, and is $f=2v_r/\lambda$, where v_r is the component of meteor velocity in the signal direction and λ is the wave-length in the portion of the ionosphere in which the meteor is traveling. If the frequency used is sufficiently high so that scattered reflections are not obtained from the skip zone, it is certain that the error in determination of meteoric velocities resulting from assuming free space λ will be less than a percent. Since $v_r=\langle PN \rangle v/r$, f becomes $2\langle PN \rangle v/r\lambda$. If we measure time from the instant that $\langle PN \rangle = 0$, $\langle PN \rangle = -vt$, and the whistle frequency is

$$f=-2v^2t/\lambda(D^2+v^2t^2)^{\frac{1}{2}}. \qquad (1)$$

A plot of this equation is shown in Fig. 5.

The maximum whistle pitch that can be obtained is evidently $2v/\lambda$, and corresponds to the condition $t=-\infty$ when the meteor is observed head on. Because of the low signal strength returned from meteors when the angle $\langle OPN \rangle$ of Fig. 4a is acute, it has been found experimentally that only the linear portion of the whistle pitch curves of Fig. 5 corresponding to t near zero is ordinarily obtained. For positive t, of course, the curve is not applicable because of the introduction of the reflection condition of Fig. 4b. The shape of the curve of Fig. 5 is determined by just the velocity v and perpendicular range D. In theory, then, experimental observation of a down whistle should enable the velocity and range to be found. Actually, since only the linear part of the whistle pitch curve is available, the information derivable from the curve is reduced to the slope df/dt measured for $t=0$. If we call this

slope fo', we have from Eq. (1)

$$fo' = -2v^2/\lambda D, \tag{2}$$

and a relation is determined between meteoric velocity and range. Fortunately, it is easy to obtain the range independently by pulse observation of bursts.

There is an interesting modification of the pulse technique which should enable ranges to be found from the same signal used to observe whistles. It involves the use of negative pulses; the continuous wave transmitter is pulsed off instead of on. With this method of operation the carrier is on 99 percent of the time so that whistles can be observed in the normal manner. Being a beat frequency resultant between the ground wave and a Doppler reflected sky wave, the whistle will vanish whenever either the ground wave or sky wave is missing. When using negative pulses at the transmitter the whistle will vanish twice; it vanishes first when the ground wave goes out, and again later when the meteor signal goes out. The separation in time between the two zeros in whistle amplitude gives the range of the meteor. When the whistle signal is of the order of magnitude of the noise the method won't work, since the pulses are unidentifiable. Negative pulses put nicks in the burst wave form which are of value for timing purposes, especially if more than one receiver is in use. In general, a separate pulse transmitter for range determination is desirable.

V. THE DETERMINATION OF DIRECTION AND POSITION FROM WHISTLE ANALYSIS

Under one rather powerful restriction it is theoretically possible to completely locate a meteor path in space by making three simultaneous whistle observations and one pulse observation at three receiving locations. The restriction is that it be possible to detect the same meteor at all three locations. Because of the aspect effect which prevents reflections from being returned when the angle $\langle PNO \rangle$ of Fig. 4a is too small, only a fraction of those meteors detected by one ground station can be detected at two others. The conditions which favor coincidental detection are high transmitted power and low operating frequency, since these conditions favor a strong reflected signal. Closely spaced ground stations increase the likelihood of coincidence, but reduce the possible accuracy of measurement. It should be noted that the probability of coincidence for whistles is much greater than for bursts.

Reflections from meteors in the general case involving a separated transmitter and receiver differ from those in the special case previously discussed in that the reflected waves do not retrace their courses. The Doppler shift, which is caused by the changing total path length from transmitter to meteor to receiver, is now composed of two parts which must be treated separately. The first path, from transmitter to meteor, is the same as that traversed by the wave reflected back to the transmitter site, so that we can immediately say that there will be a component of shift in the received frequency of just half the value given by Eq. (1). That is

$$f_{om} = -v^2 t_o/\lambda(D_o^2+v^2t_o^2)^{\frac{1}{2}}, \tag{3}$$

where the subscript o refers to the transmitter site, f_{om} is the part of the Doppler shift resulting from the variation of the transmitter-meteor path, D_o is the perpendicular distance from the transmitter to the meteor path, and t_o is time measured as zero when the meteor is at the point from which D_o is measured.

The second component of the Doppler shift results from the variation of the meteor-receiver path. Since the geometry of this variation is similar to that leading to Eq. (3), if we designate the receiving location by the subscript 1 we have

$$f_{m1} = -v^2 t_1/\lambda(D_1^2+v^2t_1^2)^{\frac{1}{2}}, \tag{4}$$

where D_1 is the perpendicular distance from the meteor path to the receiver, and t_1 is time measured as zero when the meteor is at the point from which D_1 is measured. The total Doppler shift from transmitter to receiver number one is $f_{om}+f_{m1}=f_{o1}$,

$$f_{o1} = -v^2 t_o/\lambda(D_o^2+v^2t_o^2)^{\frac{1}{2}} - v^2 t_1/\lambda(D_1^2+v^2t_1^2)^{\frac{1}{2}}. \tag{5}$$

Note that when the transmitter and receiver are not at the same location, the whistle pitch curve is the sum of two curves of the form of Fig. 5. The asymptotes representing velocity are the same for the two component curves, but the curves may differ both as to the rapidity with which they pass through zero, and as to the time at which zero frequency is reached. The complex

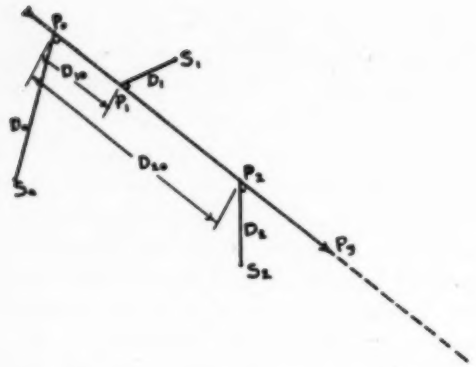


FIG. 6. The perpendiculars to a meteor trail from three ground stations.

origin of the pitch of Eq. (5) would make its analysis unavailing were it not possible to obtain the two component frequencies separately.

To find the two terms of Eq. (5) separately, it is only necessary to note that the first term, f_{om}, is just half of the value f_{oo} given by recording the whistle with the receiver adjacent to the transmitter. If we take half of the whistle pitch curve f_{oo} and subtract it from f_{o1}, we have left just f_{m1}. By making both pulse measurements and whistle measurements with the receiver at the transmitter, v and D_o are determined. From f_{m1} it is then possible to get the range D_1. Also, the difference $t_1 - t_o$ can be obtained by measuring the time interval between the instants at which f_{om} and f_{m1} go to zero. This difference in time is obviously a measure of the distance D_{10} along the meteor path between the perpendiculars D_o and D_1.

It is evident that if we install a third remote receiving location, denoted by the subscript 2, we shall be able to determine its perpendicular distance to the meteor path D_2, as well as the location of the perpendicular relative to D_o as specified by D_{20}.

We shall now proceed to show how a knowledge of the perpendicular distances D_o, D_1, and D_2, and of the path lengths between perpendiculars D_{10} and D_{20} can be used to completely specify the position of the meteor. Consider Fig. 6. Let P_g be the point where the extended meteor trail would intersect the ground. The position of the meteor will be completely specified if P_g is located and if the azimuth and incidence angle of the meteor trail are given. It is agreeable to find that the direction of arrival of the meteor is determinable merely from a knowledge of D_{10} and D_{20}. That this is so follows from the fact that stations S_o, S_1, and S_2 lie within planes through points P_o, P_1, and P_2 perpendicular to the meteor path. These planes, where they intersect the ground, generate three straight lines through the three stations, and the perpendicular to the direction of these lines gives the azimuth of the meteor path. If the meteor makes an angle ζ with the vertical, the planes through P_o and P_1 will generate straight lines on the ground having a spacing $D_{10} \sec \zeta$, since the planes themselves are spaced D_{10}. We must find these lines on the ground, because if we know them, their direction determines the meteoric azimuth, while their spacing serves to determine the incidence angle ζ.

Consult Fig. 7. We may draw a straight line L_x through stations S_o and S_2. Since the planes through P_o, P_1, and P_2 and perpendicular to the meteoric path are spaced by D_{10} and D_{20}, we must have $\langle P_x S_o \rangle / \langle S_2 S_o \rangle = D_{10}/D_{20}$, and point P_x is determined as the point on L_x intersected by the plane through P_1. But this plane through P_1 also passes through S_1, so L_1 is determined as drawn. Now L_o and L_2 are determined by the intersection with the ground of parallel planes passing through S_o and S_2, so L_o and L_2 are drawn parallel to L_1, if they are desired. We know that wherever the meteor may be, its direction is perpendicular to these lines.

Next we must measure the spacing between, say, L_o and L_2. In Fig. 7 this spacing is shown by the dotted line. We know that this spacing must be $D_{20} \sec \zeta$ because of the inclination of the planes through P_o and P_2. Dividing $D_{20} \sec \zeta$ by D_{20}, which is known, gives $\sec \zeta$, and the direc-

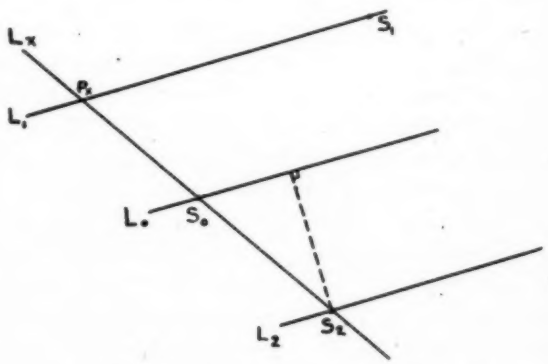


FIG. 7. Construction used to determine the direction of meteoric arrival.

tional characteristics of the meteor are completely determined.

Now that we have seen how simply the meteoric azimuth and angle of incidence were found, we must investigate means of finding the location of the point P_g, where the extended meteor trail pierces the ground. The information we have to use is, in addition to the angles just determined, knowledge of the perpendicular ranges D_o, D_1, and D_2. We know, for instance, that the meteor is somewhere in a cylinder having S_o on the axis, and with D_o as radius. The direction of the axis has been found. The projection of this cylinder on the ground will be an ellipse having minor axis D_o in the direction of line L_o, and with major axis $D_o \sec\zeta$ at right angles. There are, similarly, ellipses surrounding S_1 and S_2, with ellipticity determined by ζ, and minor axes D_1 and D_2. The three ellipses thus determined must intersect at the point P_g, since they each claim the meteor path as one of their elements.

Knowing the major and minor axes of these three ellipses, and the directions of their axes, we might now proceed to plot them on a large map of the terrain, and seek their common intersection. Rather than do this, we may note that we may transform the map, and make the ellipses into circles. The required transformation of the map may be effected by reducing all dimensions perpendicular to the family of lines L by the factor $\cos\zeta$. This is simple, since S_o, S_1, and S_2 are the only points needed on the transformed map. Dimensions parallel to the family L are left unaltered. In this transformed map the ellipse about S_o becomes a circle of radius D_o and, similarly, the ellipses about S_1 and S_2 are circles of radii D_1 and D_2. Their common intersection determines P_g in the transformed map, and by expanding by $\sec\zeta$ perpendicularly to the family L, the point P_g can be plotted in the original map.

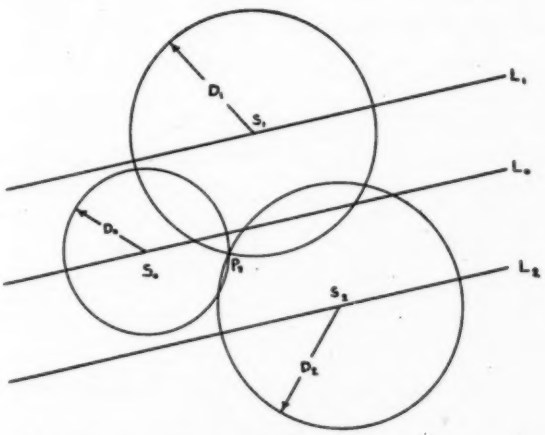


FIG. 8. Construction for locating the point of ground intersection of the extended meteor trail.

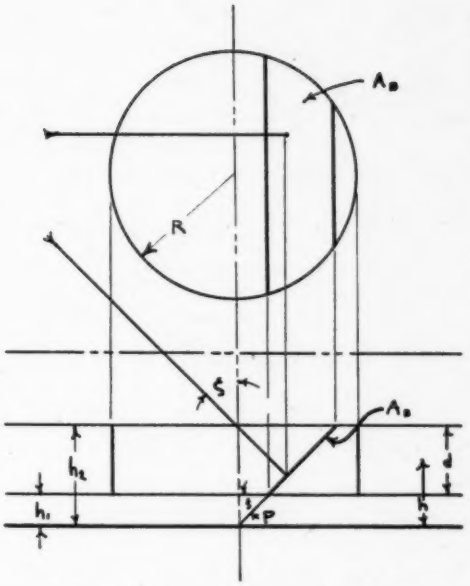


FIG. 9. The region in which bursts are produced.

Figure 8 shows the solution for P_g in the transformed plane. Note that two circles serve to locate the meteor with the exception of a possible choice between two points. The third circle resolves the ambiguity while at the same time giving an idea as to the reliability of the results.

VI. THE FRACTION OF METEORS PRODUCING BURSTS

When tallying meteors by observing bursts, allowance must be made for the fact that only a certain fraction of the meteors passing within the range of the apparatus are capable of producing reflections. The actual percentage of meteors that are geometrically located so as to be able to produce bursts depends upon the angle ζ of the meteor path with the zenith. If directive antennae are used, as is the case in making radar observations, the meteor count will also depend radically upon the direction from which the signals are being received. The reason for this state of affairs is in essence contained in Fig. 9. There it is shown

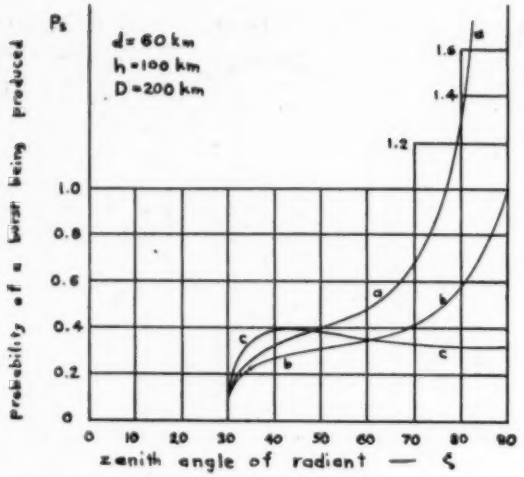


FIG. 10. The probability that a given meteor will produce a burst on three theories.

that in order to produce a burst, the reflecting point must be in a plane perpendicular to the direction of meteor travel. This plane is labeled P. Hence, for all meteors from a given radiant the bursts will be found within this plane. It is this directive property of the bursts which makes it possible to deduce the radiant of a meteor shower by radar observations, as has been done by Hey and Stewart.[4] Their curves may be quoted as an illustration of the marked dependence of the observed meteor count upon the direction of antenna orientation, and their successful radiant determinations may be cited as proof that bursts are reflected with the incident and reflected rays making supplementary angles with the meteor path.

Even when the most advantageous antenna arrangements are used, only a fraction of the observable meteors will produce bursts. The essential difficulty is that a meteor can in general only produce a sufficient degree of ionization to reflect waves when within a certain region of altitude bounded by the heights h_1 and h_2. A burst will be produced if the perpendicular to the meteor trail lies on the portion of the path within h_1, h_2. Otherwise no burst will result, and only whistle observations will detect the meteor. The calculation of the percentage of meteors capable of producing bursts must necessarily be quite inexact. In the first place, perhaps the altitudes h_1 and h_2 cannot rightly be assumed the same for all meteors. In addition, the maximum range from the observatory at which bursts may be observed cannot be assumed to be fixed. All in all, the calculation cannot be expected to be quantitatively useful, but it does illustrate the manner in which the fraction varies with the parameters of importance.

Pierce has already treated the problem under the assumption that it is equally probable for a meteor to come from any portion of the celestial sphere.[2] The basis of his treatment may be seen by reference to Fig. 9.

In order to produce a burst a meteor trail must pierce plane P between the altitudes h_1 and h_2, and it must be within a radius R in the h plane. The area on P thus defined is A_B. Pierce takes the total area upon which meteors may impinge within the radius R as πR^2, and its component in a plane parallel to P as $A_T=\pi R^2\cos\zeta$. He computes the area A_B as

$$A_B=2d(R^2-h^2\cot^2\zeta)^{\frac{1}{2}}/\sin\zeta \qquad (6)$$

if $d=h_2-h_1$ and $h=(h_2+h_1)/2$. He does not present the fraction of shower meteors from a given radiant that will produce bursts, but he does find a measure of the number of meteors from purely random directions which will fall on the areas A_B and A_T, and so computes the probability P_R that a random meteor will produce a burst. His result is

$$P_R=(2d/R^2)\cdot((R^2+h^2)^{\frac{1}{2}}-h). \qquad (7)$$

If we introduce the distance D, such that $D^2=R^2+h^2$, D is the maximum distance to a detectable meteor, measured from the observatory. We then find that a simpler representation of Pierce's formula for P_R is

$$P_R=2d/(D+h). \qquad (8)$$

As an example of the magnitudes involved, suppose $d=60$ km; $D=200$ km; $h=100$ km; these distances were used by Pierce as an illustration. Then Eq. (8) gives $P_R=0.4$. By using A_B and A_T as calculated by Pierce we may now find the probability P_S that shower meteor coming in at an angle ζ will produce a burst. By simply taking the ratio A_B/A_T we get

$$P_S=4d(1-(h/R)^2\cot^2\zeta)^{\frac{1}{2}}/\pi R\cdot\sin 2\zeta. \qquad (9a)$$

If we proceed thus, we will find, however, that the probabilities computed for large angles ζ will

be far too large, and for $\zeta = 90°$ Pierce's values give a probability of infinity, as in curve (*a*) of Fig. 10. The difficulty is that in calculating the total projected area A_T of the cylindrical region Pierce neglected the possibility that meteors might enter through the sides as well as the top. With this correction, we find $A_T = \pi R^2 \cos\zeta + 2Rd \sin\zeta$; the second term is new. We may now evaluate P_S again as

$$P_S = (4d/\pi R)(1-(h/R)^2 \cot^2\zeta)^{\frac{1}{2}}/ (\sin 2\zeta + (4d/\pi R)\sin^2\zeta), \quad (9b)$$

and curve (*b*) of Fig. 10 shows a plot of the corrected relation, again using $d = 60$ km; $h = 100$ km; and $D = 200$ km. This equation is not subject to the inconsistency contained in Pierce's analysis, but does still contain the rather dubious assumption that all meteors succeed in penetrating to the same minimum altitude regardless of their inclination. An assumption perhaps equally wrong, but erring in the other sense, is that the ionization path length is independent of the path inclination. We may modify the foregoing analysis to include this assumption by taking $d = l\cos\zeta$, if l is the ionization path length. Making this substitution in Eq. (9b), we get

$$P_S = (1-(h/R)^2 \cot^2\zeta)^{\frac{1}{2}}/ \sin\zeta[\sin\zeta + (\pi R/2l)], \quad (9c)$$

which is plotted as curve (*c*) of Fig. 10 for $l = 60$ km. It is to be expected that the variation with ζ, of the true probability that a meteor of given mass and velocity can produce a burst, will lie between curves (*b*) and (*c*) of Fig. 10.

Upon examining Fig. 10 one is led to suspect that Eq. (8), being based upon curve (*a*), is in considerable error. To obtain an idea of the extent of the inaccuracy of this equation it is possible to perform the averaging process over all angles ζ for the corrected Pierce analysis of Eq. (9b). Doing so, we find that the same number of meteors will intercept area A_B, but that the number intercepting the projection of the cylinder is greater than the number cutting its top by a factor of $1+(d/R)$. If $d = 60$ km, $R = 173$ km, $h = 100$ km, this factor is 1.346, and P_R becomes 0.297 instead of 0.400. If the analysis of Eq. (9c) were used, yet another answer would be obtained for P_R.

Equations (9) and Fig. 10 emphasize that counting meteor rates by observing bursts, as by using a radar set, is at best only an approximate way to gauge the frequency of meteoric incidence. The difficulties in accurately determining the form of Fig. 10 are precisely the difficulties of correcting a burst count to give the true meteoric frequency. Worse than that, the parameters d, R, and to a lesser extent h, cannot be determined with accuracy because of the variable nature of the incident meteors both as to size, velocity, and orientation.

VII. CONCLUSION AND ACKNOWLEDGMENT

An attempt has been made to describe the basic methods used to detect meteors by radio methods, and to indicate the nature of the data provided by some of the simpler experiments. These concepts as to the mechanism of meteoric detection have formed the basis for some more detailed investigations carried on at Stanford since the winter of 1946–7, and to be described in forthcoming papers.

With pleasure I wish to express my indebtedness to those of my associates with whom these meteoric investigations have been undertaken. Special thanks must be extended to Oswald G. Villard, Jr., Robert A. Helliwell, and William E. Evans, Jr. I should like also to thank Dr. Hugh H. Skilling, Dr. Leonard F. Fuller, and Dr. Frederick E. Terman for the stimulating interest which they have each shown in our investigations.

Aids to Analysis of Patterns Obtained in the Diffraction of Electrons by Gases

H. J. YEARIAN AND W. M. BARSS*
Department of Physics, Purdue University, Lafayette, Indiana
(Received August 22, 1947)

Three techniques have been found particularly useful in the analysis of the electron diffraction patterns produced in the scattering of electrons by gases. The first is the use of a continuously rotating shutter which is placed directly in front of the photographic plate used to record the patterns. The shutter has an angular opening proportional to the third power of the distance from the center of the pattern and serves to reduce greatly the high background intensity gradient which is characteristic of electron scattering. The second is the application of a Fourier computing scheme to the transformation from intensity curves, derived from observed patterns, to the "radial distribution function" which is of value in selecting a model for the molecule whose structure is being determined by the analysis. The third technique is the computation of patterns corresponding to various molecular models by means of a graphical synthesizer in $(\sin x)/x$. This synthesizer has advantages over the usual numerical computing schemes in that it takes account of the angular variation of atomic structure factor for electron scattering and also makes evident the contributions of scattering by specific atom pairs to the complete synthesized intensity curve.

INTRODUCTION

RESEARCH on the structure of gaseous molecules by means of electron diffraction is characterized by many procedures which are liable to subjective errors. Diffraction patterns are usually measured visually and the intensity curves on which the analysis depends are based on the visually estimated relative intensities of adjacent rings in the patterns. The methods of harmonic analysis of intensity curves, and the determination of molecular parameters by comparison of the intensities calculated for molecular models with those actually observed, are made difficult of interpretation and subject to error by the absence of more accurate intensity data.

Improved methods of obtaining more reliable intensities have been available for some time but have not come into general use. The techniques to be described have been found to give satisfactorily objective intensity data and to make full utilization of their potentialities without undue increase in the time and effort of analysis.

THE ROTATING SHUTTER

One of the characteristics of electron diffraction by gases is that the intensity of the diffraction patterns varies approximately as the inverse fourth power of the angle variable s, ($s=4\pi(\sin\theta)/\lambda$, where θ is half the angle of scattering) or of the distance, R, measured from the center of the pattern. Since the contribution of the molecular interference pattern is small compared with the monotonically varying background intensity, microphotometer traces of uncorrected photographs of patterns show no real maxima or minima, although these may be apparent under visual observation. Some approximation to a flat background may be obtained by copying the original pattern photographically, using a radial correction factor provided by a suitable shutter, but it has been shown by P. Debye[1] that such photographic modification cannot effect complete compensation. P. Debye[2] introduced a rotating shutter directly in front of the photographic plate of his diffraction camera to flatten the background on the original photograph. The angular opening of the shutter varied as R^3, or as R^4, so that the exposure time increased with radius to compensate for the decreased intensity. The R^4 shutter actually overcompensates because the variation in atomic structure factor with diffraction angle reduces the intensity of scattering at the smaller angles. P. Debye[3] later constructed shutters with

* Now at the National Research Council, Ottawa, Canada.

[1] P. Debye, Physik. Zeits. **40**, 507 (1939).
[2] P. Debye, Physik. Zeits. **40**, 66, (1939).
[3] P. Debye, Physik. Zeits **40**, 404, (1939); For other applications of the shutter (Sector) technique see: C. Finbak, O. Hassel and B. Ottar, Arch. Math. Naturv. **B44**,

angular openings designed to compensate for all variations of the intensity due to coherent atomic scattering. This method, however, requires a new shutter for each compound and every voltage at which it is desired to work.

In their electron diffraction camera the authors have used a shutter with angular opening proportional to the third power of the radius which gives this correction over the range from 4.5 to 38 mm. This gives a background sufficiently flat that the maxima and minima are real and the measured intensities may be further corrected by any desired factor. The shutter was first operated intermittently by means of a pinion and spring loaded rack. It was later modified by providing a continuous gear drive with power transmitted through the wall of the vacuum system by means of a low speed shaft operating in a Wilson seal. The shutter speed is from thirty to sixty revolutions per minute.

The high shutter speed attainable provides effectively uniform circumferential intensity for even the shortest exposures used. The effective range in angle variable, s, may be extended by recording diffraction patterns obtained for different values of accelerating voltage applied to the electron gun. Composite intensity curves are obtained from the microphotometer traces by correcting for the emulsion characteristics of the photographic plate, transforming from the radial variable R to the angle variable s, and making any necessary correction of the intensities for eccentricity of the pattern with respect to the shutter. They cover a range in s of from about 3.5 to over 25 reciprocal angstroms and permit the quantitative comparison of intensity values over a wider range than is possible with intensity curves derived by visual methods. A typical microphotometer record is shown in Fig. 1.

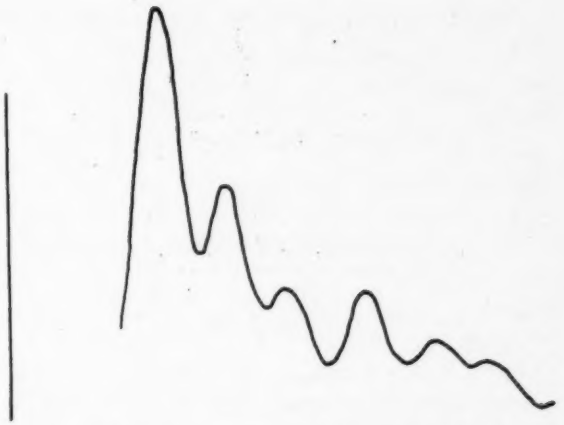


FIG. 1. The diffraction pattern of chloropicrin as photographed on Lantern Slide, Medium, plates through a shutter for which the exposure varies as R^3. In order to reduce grain variations of the plate, microphotometer records along three different radii have been superimposed and an average curve drawn.

FOURIER COMPUTATION

Although the most accurate determination of the structure of a gaseous molecule by electron diffraction requires the comparison of the observed intensity curve with curves calculated for a series of molecular models, the selection of an appropriate model may be facilitated by the application of a Fourier transformation method of the type described by Pauling and Brockway[4] to obtain a "radial distribution curve" whose peaks indicate the interatomic distances in the molecule.

The expression used to calculate the radial distribution function, $D(r)$, is

$$D(r)=k\int_0^\infty s^6 I(s)(\sin sr/sr)ds. \qquad (1)$$

where $I(s)$ is the intensity of molecular coherent scattering observed at the angle designated by s. In their work Pauling and Brockway approximated the integral by a summation

$$D(r)\doteq\sum_p I_p(\sin s_p r/s_p r) \qquad (2)$$

where I_p is the estimated intensity of the pth maximum or minimum and s_p is the corresponding value of s. (In the original work only the maxima were used.) Where an intensity curve is available it is possible to make a closer approximation to the integral and obtain greater accuracy in the radial distribution curve. Since the intensities cannot be observed over the full range of s required for the integration, it is advisable to modify the intensity curve by using the factor $\exp(-as^2)$ introduced by Degard,[5] which has the effect of reducing the relative contribution of the outermost intensity peaks.

No. 13, (1941); C. Finbak, Avh. norske Vidensk Akad. M.N. K1, No. 7, (1941); H. J. Yearian, Phys. Rev. 59, 926 (1941); S. H. Bauer, Phys. Rev. 64, 316, (1943).

[4] L. Pauling and L. O. Brockway, J. Am. Chem. Soc. 57, 2684 (1935).

[5] C. Degard, Bull. Soc. Roy. Sci. Liége 12, 383, (1937).

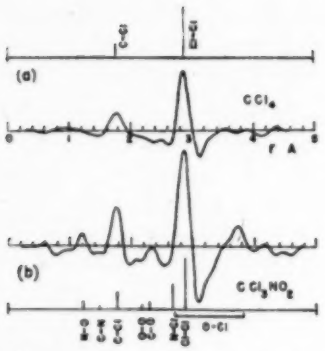


FIG. 2. Radial distribution curves obtained by Fourier analysis of diffraction patterns using the shutter technique. (*a*) CCl_4, (*b*) CCl_3NO_2.

To evaluate integral (1) making full use of a complete intensity curve would require a tedious numerical integration for each desired point on the $D(r)$ curve, unless a harmonic analyser is available, or some numerical form of Fourier analysis is used. We have found the Fourier computation scheme of Danielson and Lanczos[6] is very convenient for this purpose. All computation can be done in a relatively short time on a standard calculating machine. An analysis based on 36 measured values of $s^5I(s)$ in Eq. (1) (modified by a Degard factor) covering a range in s of $24A^{-1}$, will give 36 values of $rD(r)$ over a range in r of 4.7A. Other points at critical parts of the curve may be interpolated at will.[7]

Figure 2(a) shows the curve $r^2D(r)$ obtained in this way for CCl_4. The peaks corresponding to the accepted interatomic distances shown are well marked.[8] Figure 2(b) gives the radial distribution curve of CCl_3NO_2 which is closely related to CCl_4 in that one of the tetrahedrally arranged chlorine atoms is replaced by the nitro group.* The C−Cl and Cl−Cl distances remain with little change, but new peaks are present corresponding to the N−O, C−O and perhaps the O−O distances indicated (taken from nitromethane). The very weak peak expected for C−N is not observed, and N−Cl is not resolved. No definite peaks are evident for the various O−Cl distances, but these are grouped within the range shown with values depending upon the orientation of the nitro group, or will vary continuously within these limits if the group freely rotates about the C−N bond.

[6] G. C. Danielson and C. Lanczos, J. Frank. Inst. **233**, 365, 435 (1942).

[7] For other methods of evaluating the integral see: J. Walter and J. Y. Beach, J. Chem. Phys. **8**, 601 (1940); P. Shaffer, V. Schomaker and L. Pauling, J. Chem. Phys. **14**, 659 (1946).

[8] The tendency for minima on either side of prominent maxima is a real effect since the $r^2D(r)$ defined by (1) is not a distribution of atomic separations but represents an integral of the product of the charges situated in volume elements separated by distance r. When r is equal to the mean value of an interatomic distance, electron-electron and nuclear-nuclear contributions are predominate and give a positive peak, but for distances slightly different from this, electron-nuclear contributions produce a negative value. (P. Debye and M. Pirenne Ann. d. Physik **33**, 617, (1938).) For a molecule of one kind of atoms only, a suitable correction of the intensities can be made which will give a true atomic radial distribution. A similar approximate correction in the general case will yield a distribution having a physical meaning which is somewhat uncertain and has not been made here. (See C. Degard, Bull. Soc. Roy. Liége, **13**, 117 (1938); V. Schomaker, Thesis, California Institute of Technology, 1938; P. Debye, J. Chem. Phys. **9**, 55 (1941); R. Spurr and V. Schomaker, J. Am. Chem. Soc. **64**, 2693 (1942); and reference 7 above.)

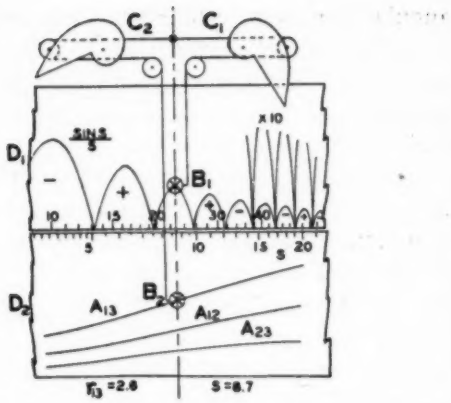


FIG. 3. Schematic arrangement of synthesizer as set to calculate $A_{ij}(\sin sr_{ij})/(sr_{ij})$ for $i=1$, $j=3$, $s=8.7$, $r_{13}=2.6$. D_1 and D_2 are drums carrying curves $(\sin s)/s$ and A_{ij} respectively, C_1 and C_2 exponential cams controlling the cross-hairs B_1 and B_2 and driving the multiplying tapes (not shown).

GRAPHICAL SYNTHESIZER IN SINX/X

A variety of methods have been used in the synthesis of diffraction intensity curves for different proposed molecular models. All involve the evaluation, at a large number of s values, of a summation of the form

$$I(s)=\sum_{ij}{}'E_i(s)E_j(s)(\sin sr_{ij})/sr_{ij} \qquad (3)$$

where $I(s)$ is the intensity of molecular coherent scattering at the angle given by s, E_i is the atomic structure factor for electrons and r_{ij} is the separation of the ith and jth atoms in the molecule. The summation is taken over all values of i and

* A detailed discussion of the structure of this molecule will be published elsewhere (W.M.B.).

j, $i \neq j$. Equation (3) may be re-written as

$$s^4 I(s) = \sum_{ij}' A_{ij} (\sin sr_{ij})/sr_{ij} \qquad (4)$$

where $A_{ij} = (Z_i - f_i)(Z_j - f_j)$, Z is the atomic number and f is the atomic structure factor for x-rays, whose dependence on s is given in the tables of James and Brindley[9] and in graphical form by Buerger.[10]

Special tables of $(\sin x)/x$ have been prepared in forms suitable for computation of the function for a selection of r-values, but such computations can be made conveniently only when the factor $z_i z_j$ is used as an approximation to the correct structure factors. The use of this approximation results in the calculated intensities being too high at small values of s. With the more detailed intensity curves obtained by the shutter technique, the complete formula should be used. To take account of the variation of structure factor with diffraction angle one of the writers[11] constructed a synthesizer consisting of graphs of the functions $(\sin sr_{ij})/sr_{ij}$ and A_{ij} on which their ordinates could be measured at desired values of s and their product calculated.

The mechanical arrangement of the synthesizer has been modified as shown in Figs. 3 and 4. Referring to the schematic diagram of Fig. 3, the function $(\sin s)/s$ is plotted on the cylindrical surface of a drum, D_1, and the amplitude functions $A_{ij}(s)$ are plotted for each atom-pair on a similar drum, D_2, mounted on the same axial shaft. All curves are plotted against $\log(s)$ on a scale such that one decimal cycle in s covers one meter. The values of sr_{ij} and s are shown by similar scales along the adjacent edges of the two drums, the logarithmic scale permitting the selection of a desired r_{ij} value by the appropriate setting of the relative angular displacement of the drums. Readings are taken along a line parallel to the common drum axis by means of cross-hairs mounted in two small carriages, B_1 and B_2, which ride on a track close to the surface of the drums.

Multiplication of the ordinates $\sin(sr_{ij})/(sr_{ij})$ and $A_{ij}(s)$, read under the cross-hairs, is obtained by moving the cross-hair carriages by means of the cams C_1 and C_2 arranged so that the angular displacement of a cam is proportional to the logarithm of the displacement of the corresponding cross-hair. Mounted on the parallel cam shafts are cylindrical pulleys carrying two parallel metal tapes, not shown in Fig. 3 but visible in Fig. 4. The tapes, on which logarithmic scales are engraved, move with the two cams in such a way that their relative displacement is proportional to the sum of the logarithms of the ordinates measured by the two cross-hairs. Thus the product of the ordinates may be read directly from the tapes as from a slide rule.

Extra curves, $10(\sin s)/s$, are drawn to permit readings to be taken effectively nearer the zero axis, the product indicated by the tapes being divided by the factor 10. Also, since the cross-hairs cannot cross the axis, the negative portions of the function $(\sin s)/s$ are plotted, inverted, above the axis and drawn in red to distinguish them from the positive portions.

As shown in Fig. 4, the drums and carriage tracks are mounted on an inclined table so that the cross-hairs may be seen easily at all positions along the drums. The cross-hairs are also illuminated at normal incidence to reduce parallax errors. The tapes carrying the scales on which the products are indicated are readily visible at

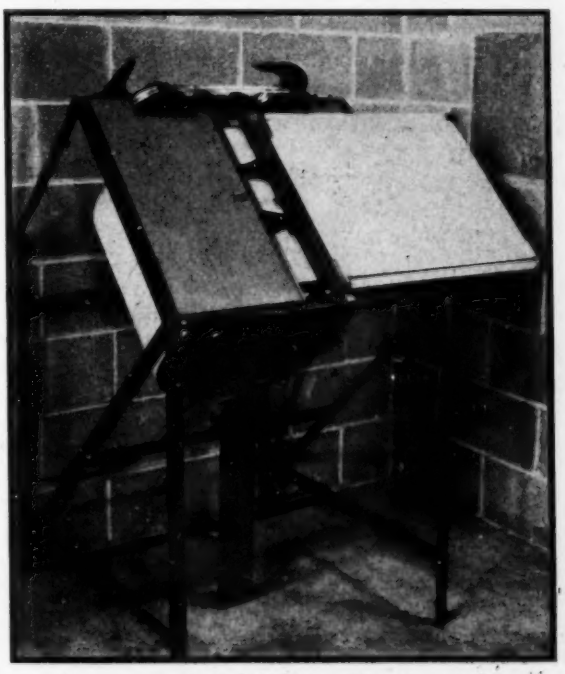


FIG. 4. View of completed synthesizer.

[9] R. W. James and G. W. Brindley, Phil. Mag. 12, 104 (1931).
[10] M. J. Buerger, *X-ray Crystallography* (John Wiley and Sons, Inc., New York, New York, 1942), p. 52.
[11] H. J. Yearian, Phys. Rev. 59, 926 (1941).

the top of the frame, and the controls for the cams are placed for convenient operation by hand. A locking mechanism on the drums is provided whereby stops may be pre-set for desired values of r, so that the corresponding relative angular displacements of the drums may be obtained quickly and accurately. A drawing board and roll of graph paper are mounted on the instrument so that the synthesized intensity values may be plotted point by point as they are obtained.

It has been found most convenient to proceed by setting the drums for an r-value corresponding to an interatomic separation, r_{ij}, reading the products $A_{ij}(\sin sr_{ij})/sr_{ij}$ for a series of s-values and plotting them as a curve $s^4I_{ij}(s)$. As shown in Eq. (4), the desired intensity curve, $s^4I(s)$, is the algebraic sum of all such component curves, $s^4I_{ij}(s)$, taken for the several values of r_{ij}. Drawing the component curves, instead of performing the summation at each value of s, gives a clear indication of the intensity contributions made by each atom pair and enables an intelligent estimate to be made of which parameters r_{ij}, if any, should be changed to improve agreement with the observed intensity curve. The "observed" intensity referred to is also of the form $s^4I(s)$, where $I(s)$ is the true electron intensity, since the rotating shutter already described gives a factor of s^3 and another power of s is readily included while making the other corrections in changing the abscissa from the ring radius, R, to the angle variable, s.

The synthesizer may also be used numerically to integrate Eq. (1). The observed values of $s^6I(s)$ (including a Degard factor if desired) are plotted on the lower drum instead of the coefficients A_{ij}, and the drum displacement is set for a value of r at which a point on the $D(r)$ curve is desired. The values of the integrand at equally spaced values of s may then be read from the tapes and the sum computed. For an analysis covering a complete range of r we find the Fourier method superior, but in cases where only a limited range of r values is of interest, this integration scheme is advantageous.

CONCLUSION

Methods have been described for the derivation of electron diffraction intensity curves which make possible the rigorous application of a Fourier analysis to obtain a radial distribution function of increased usefulness. An instrument has also been described which will aid in the synthesis of intensity curves with an accuracy commensurate with the accuracy now attainable in the observation and measurement of diffraction patterns.

Some Observations on the Formvar Replica Method, with a Note on Glass Surfaces

BEATRICE M. DEACON, S. G. ELLIS,* W. G. CROSS,** AND R. S. SENNETT
Department of Physics, University of Toronto, Canada

(Received January 15, 1948)

A study of the capabilities and limitations of the Formvar replica method has been made with more particular reference to shadow cast Formvar replicas. Observations have been made on the dependence of the deduced surface structure on the method of replication of the surface and the method of shadow casting the replica. The most suitable thickness of Formvar and evaporated metal film has been studied for various surfaces. Other replica methods have been summarized and compared with Formvar replica methods.

The structure of glass surfaces, both polished and fractured, is briefly discussed on the basis of observations made from Formvar replicas.

INTRODUCTION

IN 1945 this laboratory was asked to make electron microscope studies of the sleeks on polished optical components manufactured by Research Enterprises, Ltd., Leaside, Ontario. Sleeks are fine troughs in the surface of glass sometimes produced during polishing. They are without longitudinal variations in structure and are to be distinguished from scratches, which

* Now at RCA Laboratories, Princeton, New Jersey.
** Now at the Lyman Laboratory of Physics, Harvard University.

show evidence of cracking and grinding when examined in the light microscope.

From the experiments then made it became apparent that mechanically polished and, more particularly, fire polished glass surfaces containing sleeks were very suitable test objects for the study of the capabilities and limitations of the Formvar replica method. This was due to the relative simplicity of the surface structure, the ease with which the surface could be cleaned, the ease with which the Formvar films could be removed from the surface, and the fact that with fire polished glass microscope slides one could examine and identify specific areas of the glass quite readily under dark field illumination with a light microscope.

In the intervening period, since this work was commenced, a number of isolated experiments with Formvar replicas have been performed by the authors. The purpose of this paper is to coordinate and report the results of these studies, with more particular reference to the properties of the replicas than to the nature of the glass surfaces.

THE PREPARATION OF THE SURFACES

The mechanically polished optical surfaces were not treated in any way before the replicas were made, except for the cleaning process described below. In the case of the fire polished microscope slides two families of sleeks, at an angle to each other, were made on their surfaces by lightly rubbing them with fine emery cloth.

The microscope slides were generally cleaned by scrubbing them gently with a camel hair brush while they were immersed in a two percent Dupinol solution. They were then rinsed in distilled water and spun in a centrifuge until dry. In some cases surfaces were cleaned by stripping thick Formvar films from them with Scotch Tape. When such surfaces were flamed,[1] in an attempt to secure still greater cleanliness, it became impossible to remove the Formvar replicas from them.

No etching was performed on these surfaces, and the replicas were made from them very shortly after they had been cleaned.

[1] Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1945), p. 166.

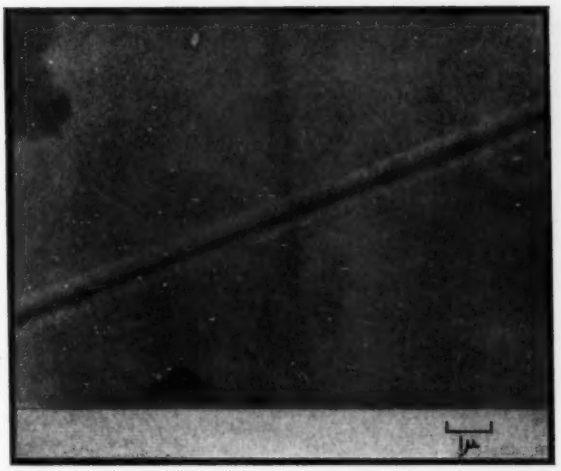


FIG. 1. Formvar Replica of surface of glass slide showing intersection of sleeks.

FORMVAR REPLICAS

The general procedures for making Formvar replicas have been well described elsewhere.[2] In this work the replicas were made from Formvar 15–95 dissolved in ethylene dichloride, at strengths varying from 0.1 to 2.0 g per 100 ml of solvent (0.1 percent to 2.0 percent solutions in the nomenclature employed hereinafter).

The cleaned glass slides were examined under dark field illumination at ×100 in the light microscope, and when a suitable area with sleeks had been found a crayon circle was made under it on the reverse side of the glass. The Formvar film was then produced on the glass, a stainless steel mesh placed on it over the crayon circle, and the replica and mesh removed together either by floating them off on water or by stripping with Scotch Tape in the case of the thicker (1 and 2 percent) replicas.

It was found necessary to use an objective diaphragm in the observation and photography of these replicas in the electron microscope. The thinner replicas gave better contrast with the smaller sleeks so that most of this work was done with 0.1 percent and 0.2 percent Formvar replicas. The thinner replicas would of course be more prone to distortion during and after removal from the glass.

Figure 1 is a micrograph of such a replica. A point of immediate interest is that the dark

[2] V. J. Schaefer and D. Harker, J. App. Phys. **13**, 427 (1942).

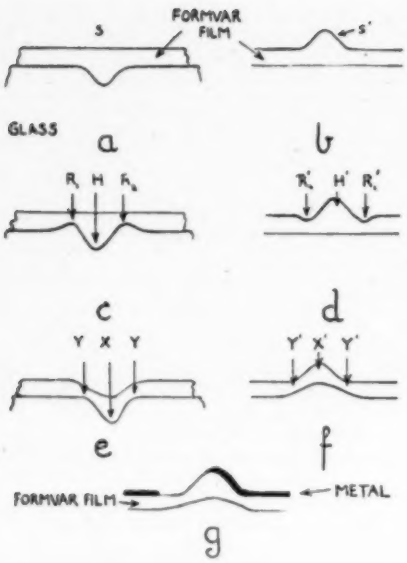


FIG. 2. Diagram showing possible contours of Formvar film in relation to sleek contour.

line, corresponding to the trough of the deeper sleek, is bounded by lines lighter than the background. There are two possible reasons for this. In Fig. 2c, if the sleek H is bounded by ridges R_1 and R_2, the replica may have the form shown in Fig. 2d, H' giving the dark line and R_1' and R_2' giving the lighter lines since the replica is thinner here than elsewhere. Alternatively the sleek Fig. 2e may not be bounded by ridges but the Formvar may be thinner at Y because of the action of surface tension at a convex surface and thicker at X where the surface is concave. The resulting replica, Fig. 2f, would give the same image in the microscope as that of Fig. 2d, since the electron scattering is a function of film thickness in the direction of the electron beam, not film shape.

This ambiguity in the Formvar replica can be removed by shadow casting the surface of the replica that was in contact with the glass (Fig. 2g).

SHADOW CAST FORMVAR REPLICAS

Replicas produced by the above method were shadow cast with chromium to produce shadow length to object height ratios of three or four.

Figure 3 is a micrograph of such a replica shadow cast with about 80A of chromium. The micrographs of shadow casts in this paper have been printed from intermediate transparent positives to give dark shadows. The shadows are shown, wherever possible, directed downwards in the figures in order to enhance the appearance of relief. The increase in contrast is evident. There is little evidence of ridges beside the sleeks here, though in Fig. 4 there is some evidence of this effect. Comparisons of ordinary and shadow cast replicas lead to the conclusion that surface tension does play a part in determining the thickness of the film in the vicinity of irregularities in the substrate. This effect contributes to the poor contrast and resolution shown in non-

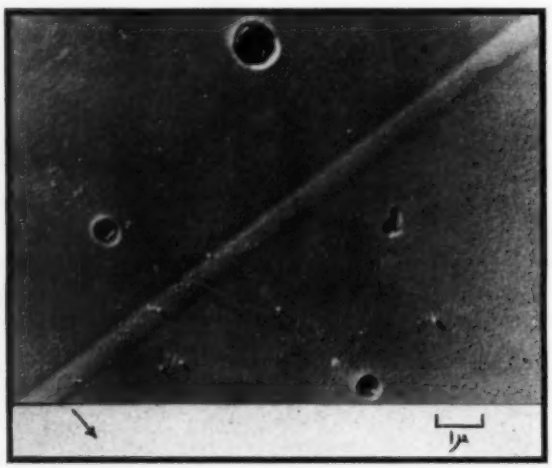


FIG. 3. Shadow cast Formvar replica of sleek in glass surface.

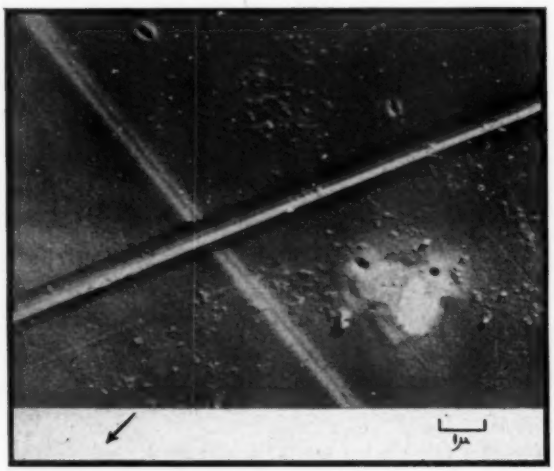
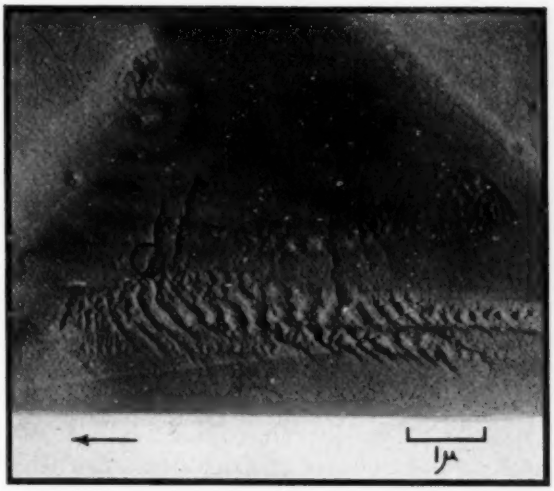
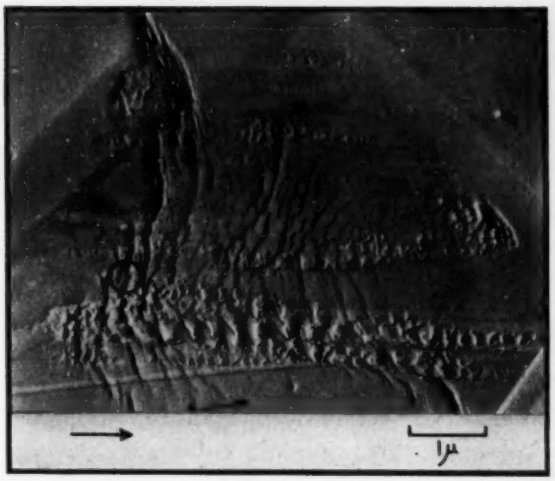


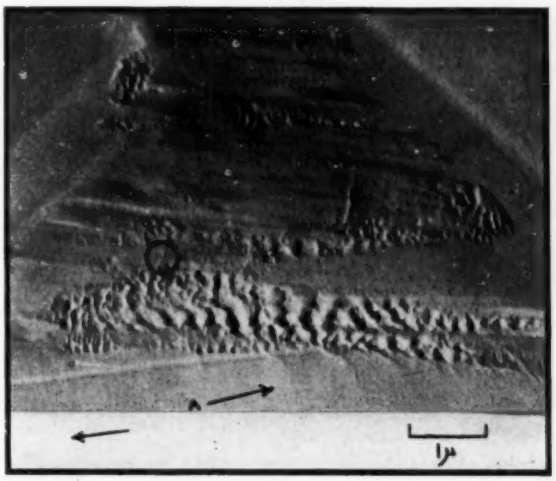
FIG. 4. Shadow cast Formvar replica showing intersecting sleeks.



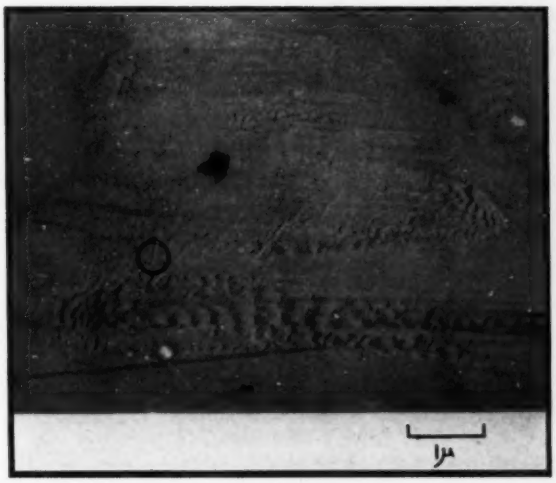
5



6



7



8

FIGS. 5–8. Different Formvar replicas of the same area of the glass surface: 5, 6, 7—shadow cast in directions indicated by arrows. 8—not shadow cast.

shadow cast replicas, and has been discussed by Heidenreich and Matheson.[3]

Experiments were next performed to test the reproducibility of the method. Two mutually perpendicular lines, between 0.2 and 1.0 micron in width, were drawn with a diamond to intersect at a region on a microscope slide which was roughened and contained numerous sleeks. The intersection served to locate and identify this region when the Formvar replica was being taken, since under oblique illumination the lines could be seen with the unaided eye. The replicas of these lines could also be identified on the intermediate screen of the electron microscope. The slide was cleaned by stripping 1.0 percent Formvar films from it. Then successive 0.5 percent replicas were stripped from the slide and most were shadow cast with chromium in thicknesses varying from 30 to 100A. In approximately one-fifth of the mounts the desired area was obtained over a mesh opening. Figures 5 to 8 show some of the results which will be analyzed below. The arrows show the direction in which the incident chromium was moving. It had become apparent in the course of this work that, with shadow casting, thicker replicas could

[3] R. D. Heidenreich and L. A. Matheson, J. App. Phys. 15, 423 (1944).

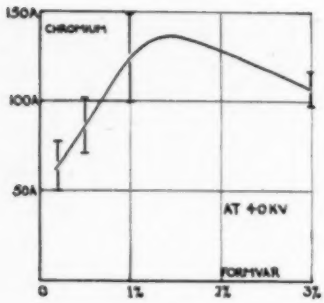


FIG. 9. Graph showing optimum thickness of chromium for given thickness of Formvar.

be used and still show adequate contrast and detail. Since this would also minimize distortion in the replicas, experiments were undertaken to discover the influence on the quality of the micrographs of varying the thickness of the replicas and also the thickness of the chromium independently.

To this end groups of replicas were made from a fractured glass rod with 0.2, 0.5, 1.0, and 3.0 percent Formvar solutions. Individual members of each group were shadow cast with approximately 70, 100, 150, and 220A of chromium, the thickness being measured normal to the plane of the replica. Micrographs were taken at an electron energy of 40 kev. Comparing the replicas with a common thickness of Formvar, the range of chromium thicknesses giving the best quality in the micrographs was judged. This most suitable range of thicknesses is plotted against the Formvar thickness in Fig. 9. If too little chromium is used the micrographs approach the poor contrast and resolution of the micrographs from non-shadow cast replicas. If too much chromium is used the finer detail in the replica is obliterated and the number of electrons inelastically scattered becomes too great for successful microscopy. With Formvar strengths greater than 1.5 percent the inelastic scattering due to the Formvar itself becomes so great that only lesser thicknesses of chromium can be tolerated.

CONCLUSIONS

It may be concluded from the foregoing that if one wishes to show fine structure in a surface then a replica made from a 0.1 or 0.2 percent Formvar solution, and shadow cast with about 100A of chromium, will be suitable. For larger structures, and in cases where replica distortion may be expected, a one or two percent Formvar replica with 150 to 200A of chromium should be used. These thicknesses may be revised upwards when working with electron speeds in excess of 40 kev.

If, by the least resolvable distance in a replica we mean the size of the smallest surface structure whose dimensions can be measured with an error which can be relied upon to be less than, say, 50 percent, then we can hope to give only a rough order of magnitude for such a quantity. With the replica method described above, the least resolvable distance will be of the order 200A to 500A for widths and 100A for heights or depths, and is imposed by the molecular size of the Formvar which produces the pebbly or corded structure to be seen in Figs. 3 and 4. The replicas will *detect* smaller dimensions than these, of the order 100A in width and 50A or less in height, if the chromium is deposited very obliquely in the proper thickness and the structures are not in the shadow of larger irregularities.

From Figs. 5 to 8 we conclude that the non-shadow cast replica reveals, in some part, detail which the shadow cast replicas do not; the structure, however, is more difficult to interpret. A considerable amount of the structure in Fig. 6 is due to wrinkling in the Formvar film. That some of the wrinkling is due to structure in the glass is shown by the narrow "ditch" in the circles. While in Fig. 5 there is no wrinkle near the ditch, in Fig. 6 a wrinkle follows the ditch and then continues on by itself. This effect might be expected to be more troublesome with rougher surfaces. It can be reduced by using thicker Formvar replicas. Williams and Wyckoff show micrographs that contain no wrinkles, even when the replicas are taken from comparatively rough metal surfaces. One should also notice that the sleek *A* of Figs. 6–8 is almost entirely missing from Fig. 5. The apparent variations in the width of sleek *A* are interesting and in accord with what has been said above about the influence of surface tension on the formation of the replicas. In Fig. 8, and Fig. 7 in which the replica is too lightly shadow cast, the sleek *A* appears to be about 500A in width, whereas in Fig. 6 it appears to be 150A in width. This latter figure is, we believe, nearer to the true value.

The appearance of the micrograph is seen to depend very much on the direction of shadow casting. This is not surprising, and follows from the same reason that the appearance of an aerial photograph will change with the elevation of the sun when it is taken. If one works back from the shadows to the form of the replica and then from the form of the replica to the form of the surface from which it was produced, very little of the result will have been influenced by the direction in which the shadow cast was made.

Finally, it should be emphasized that it would be very unwise to try to generalize much further on the basis of these experiments. The best replica technique for one surface may not be the best for another. The above discussion should serve to point out some of the major difficulties and limitations to be encountered in using shadow cast Formvar replicas.

OTHER REPLICA METHODS

Wyckoff[4] has classified replica methods as those producing molecular replicas and those producing atomic replicas. The purpose of this section is to review briefly some of the other replica methods for the sake of completeness.

Molecular Replicas

If the replica method employs at some stage a replica made of a plastic such as polystyrene or Formvar, then the least resolvable distance with the replica will be limited, at the best, by the molecular size of the plastic, as stated above. Such a replica is called a molecular replica, even if the plastic is used only in an intermediate stage of the process.

The molecular replica methods which have been devised include the following: silver-collodion,[5] lacquer,[6] Formvar,[2] shadow cast Formvar,[7] polystyrene-silica,[8,9] and methyl methacrylate-silica.[10] Judging by the published pictures the shadow cast Formvar and polystyrene-silica methods produce the best results and they have roughly equal resolving powers. We regret that we have been unable to make an experimental comparison of shadow cast Formvar and silica replicas from the same surface. We think, though, that their respective virtues and faults may be summarized in the following way.

Silica replicas are less likely to be distorted than Formvar replicas, but they are more difficult to prepare in many cases and also are less readily interpreted. It is likely that only the first layers of silica migrate on polystyrene so that a shadow casting effect is obtained. However, doubt as to the variation in thickness of the silica replicas makes it advisable to use stereoscopic methods of interpreting the structure, so that for surfaces such as the glass surface with sleeks the results can be obtained more quickly with shadow cast Formvar replicas.

A variation on the above methods which can be used if it proves to be difficult to remove thin Formvar films from the surface to be studied is to make thicker plastic replicas with Formvar or polystyrene—either from solutions of the plastic or with a press in the case of polystyrene. The replica is then shadow cast obliquely with chromium and the chromium backed by evaporating another film normally onto the plastic. This backing film should be of low density material such as beryllium or aluminum. The plastic is then dissolved away and the all-metal replica examined in the microscope.

Atomic Replicas

The limitations on resolution imposed by the molecular size of the plastic were circumvented by Williams and Wyckoff[11] in the replica transfer method in which the surface, or small particles on a smooth surface, are shadow cast and then the shadow cast is stripped from the substrate with a plastic film. It is desirable here that the metal used shall be evaporated to such a thickness that it does not spoil the resolution by migrating into particles bigger than 10 or 20A in diameter. As a general rule the metals which migrate least (e.g., chromium) are those which

[4] R. W. G. Wyckoff, Meeting of the Electron Microscope Society of America at Pittsburgh, 1946.
[5] V. K. Zworykin and E. G. Ramberg, J. App. Phys. **12**, 692 (1941).
[6] E. Golz, Zeits. f. tech. Physik **1**, 8 (1943).
[7] L. Thomassen, R. C. Williams, and R. W. G. Wyckoff, Rev. Sci. Inst. **16**, 155 (1945).
[8] R. D. Heidenreich and V. G. Peck, J. App. Phys. **14**, 23 (1943).
[9] C. H. Gerould, J. App. Phys. **18**, 333 (1947).
[10] A. F. Brown and W. M. Jones, Nature **159**, 635 (1947).
[11] R. C. Williams and R. W. G. Wyckoff, Science **101**, 594 (1945); Nature **156**, 68 (1945).

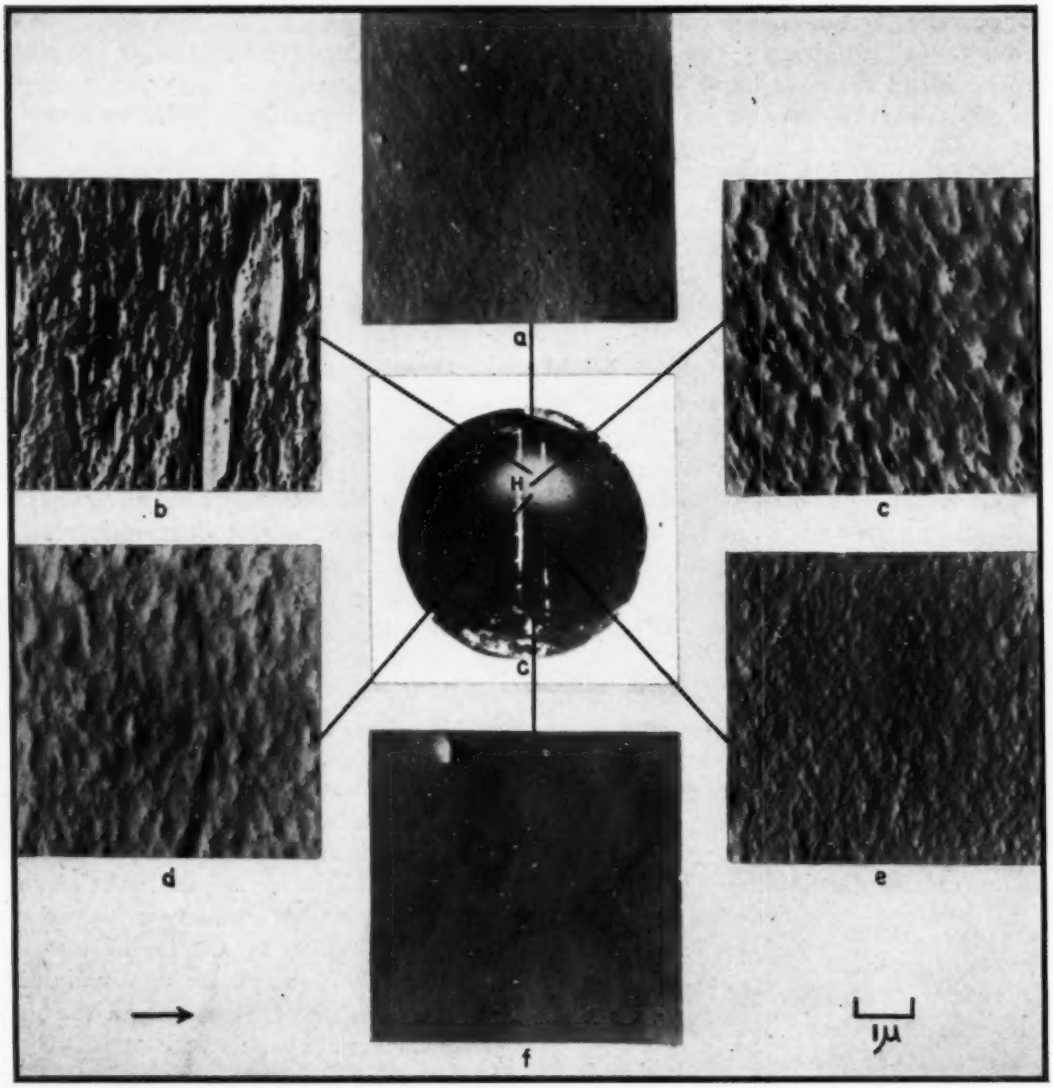


FIG. 10. Center a, b, c, d, e, f—shadow cast—photograph of surface of fractured glass rod ×1.5. Formvar replicas of surface of rod from positions indicated.

adhere most strongly to the substrate and are most difficult to strip from it. Williams and Wyckoff compromise by using gold in thicknesses less than 10A and strip it from the glass very soon after the evaporation. Since the major part of the electron scattering is done by the metal, the molecular structure of plastic film is, in this case, not a limitation. In this sense the oxide film method of Mahl[12] may also be considered to yield an atomic replica though its applicability is strictly limited and its resolving power does not appear to be much greater than that of the molecular replica methods.[3] Heidenreich and Shockley give 50A as the limit of resolution.[13]

The atomic replica process could also be used in those cases where the substrate could be dissolved away without attacking the shadow cast and backing material. From this it will be seen that replica methods in general are capable of considerable variation and modification to suit special applications.

[12] H. Mahl, Naturwiss. **30**, 207 (1942).

[13] R. D. Heidenrich and W. Shockley, J. App. Phys. **18**, 1029 (1947).

Glass Surfaces

Our conclusions as to the nature of the glass surfaces that have been studied during the course of this work may be stated briefly.

The glass polishing process has been in the past the subject of some controversy. It has been variously maintained that the fundamental action of polishing is one of grinding,[14] surface flow,[15] local heating,[16] and adsorption with subsidiary chemical reactions.[17] Replica studies of glass and other ceramics using both lacquer and aluminum films were made by Golz,[6] and our work began as a repetition of his. The conclusions we state below are in agreement with his own.

Fire polished glass surfaces show considerable areas in which the local irregularities of height are less than 100A. Sleeks made in these surfaces can be so fine that we have been unable to establish any lower limit to their size. When one sleek crosses another it "wipes out" the first one. These sleeks show no signs of chipping or gouging, so that Newton's grinding hypothesis of polishing is not substantiated. Other than to say that the polishing process is a molecular one, we are unable to decide experimentally between the other hypotheses and are inclined to feel that some of the distinctions between these hypotheses are labored.

Investigation of the process of fracture in glass has also led to the advancement of several theories.[18] In this regard a study of the structure at various regions on a fractured glass surface is of considerable interest.

If a disannealed glass rod, with a scratch around the periphery, is broken, a large mirror surface is obtained with a circular "matte" region (hackle) near the center of the rod. This can be seen in the photograph of such a surface (actual diameter 1 inch) in the center picture of Fig. 10. Light microscope observations show a continual decrease in the structure size down to the limit of resolution of the light microscope, from the hackle region *H* towards the center of fracture *C*. Electron microscope observations of replicas of the surface show a continuing decrease in the size of structure, down to the limit of resolution of the method used.

In order to study the variation in structure, 1 percent Formvar replicas were prepared from each of 9 regions. These were defined by the scratches on the fractured surface, as shown in the center of Fig. 10. The replicas were shadow cast with chromium (100A). From these, a set of 54 micrographs, of regions equally spaced across the rod, was obtained. The gradation in structure size, at various indicated positions, is shown in Fig. 10, from 10a, the position furthest from the center of fracture where the structure is comparatively fine, through 10c, where the structure is very coarse, to 10f, the position nearest the center of fracture, where the structure is the finest observed.

In all the regions studied, the individual structures show a similar shape and are oriented in such a way that their long dimension is radial

FIG. 11. Shadow cast Formvar replica of the hackle region of surface of fractured glass rod.

[14] I. Newton, *Optiks* (London, 1730), fourth edition, p. 240.
[15] J. W. French, Trans. Opt. Soc. **17**, 24 (1916).
[16] von T. Klennon Mark and U. A. W. Sinkal, Naturwiss. **11/13**, 12, 3, 143 (1943).
[17] I. V. Grebenschikor, Socialistichiskaya Reconstructeiya 1. Nauka **2**, 22 (1935).
[18] F. W. Preston, J. App. Phys. **13**, 623 (1942).

from the center of fracture. The characteristic structure is seen well in the micrograph of a 0.1 percent Formvar replica of the hackle region (Fig. 11).

Within the limit of resolution of the shadow cast Formvar replica method, no lower limit to the size of these structures has been observed. This suggests the possibility of the presence of very fine structure in fractured surfaces even in the immediate neighborhood of the center of fracture.[19]

DISCUSSION

Since the replica technique to be employed in any given problem must depend on the nature of the surface to be replicated as well as on the character of the information sought, it is undesirable to attempt to draw more general conclusions from these studies than those stated above. We have here dealt only with the factors of resolution, reproducibility, and interpretation in Formvar replicas with a few generalizations to other types of replicas.

An equally important field of study, particularly in metallurgical applications of the replica technique, is that of etchants. A metal surface is commonly polished to render it reasonably plane, before attempts are made to replicate it. The polishing will, in general, form a layer of amorphous metal on the specimen. It may be worth adding that smooth sleeks in an apparently amorphous layer have been observed even on large magnesium oxide crystals. If information is required about the bulk material, such amorphous layers must be etched away. The structure of the substrate is then made evident by the production of variations in the surface due to the differing rates at which the etchant will attack different crystalographic planes of the material if it is an element, or different components if it is an alloy or other complex solid. The work of Heidenreich, Gerould, and McNulty[20] on aluminum-magnesium alloys provides a good example of the use of etchants and contains a discussion of the factors involved in their choice. It should be noticed that in electron microscopy, unlike light microscopy, the specimens to be examined do not have to be prepared with very level surfaces. For a resolving power of 50A the depth of focus of the electron microscope is approximately ten to a hundred times greater than that of the light microscope. This indicates that with the evolution of the requisite replica techniques much rougher surfaces should be amenable to study in the electron microscope.

In conclusion it is worth stressing that the use of replica techniques with the electron microscope is a field of research that is as yet relatively unexplored. There are three main methods by which the electron microscope study of massive bodies may be carried out. These are fragmentation, sectioning, and replication. While the replica method is often considered to be restricted to surface studies, it will be seen that by deep etching, or by cutting and etching, information about the body morphology of a specimen can be obtained with replica methods. Applications have, to the present, been largely restricted to the study of hard inanimate bodies, although modifications of the method have been suggested to extend its use to the study of biological specimens such as plant or animal tissues.[21,22] It seems reasonable to expect a more rapid development in this application of the replica method.

ACKNOWLEDGMENTS

One of us (S.G.E.) wishes to acknowledge the award of a fellowship by the National Research Council of Canada during the year 1945–46, and another (R.S.S.) financial assistance from the Dentists' Supply Company of New York. It is a pleasure for all of us to record our indebtedness to Professor E. F. Burton for encouragement during the course of this work.

[19] E. F. Poncelet, Colloid. Chem. **6**, 77 (1945).
[20] R. D. Heidenreich, C. H. Gerould, and R. E. McNulty, A.I.M.M.E. Tech. Pub. No. 1979.
[21] R. B. Barnes, C. S. Burton, and R. G. Scott, J. App. Phys. **16**, 739 (1945).
[22] F. Heinmets and J. T. Quynn, Meeting E.M.S.A., Philadelphia, 1947.

Pseudostructures in Electron Microscope Specimens

JOHN H. L. WATSON
Medical Research Institute, Henry Ford Hospital, Detroit 2, Michigan
(Received January 5, 1948)

Pseudostructures, which are not real properties of electron microscope specimens but are introduced by the action of the electron beam, are described. Examples are given from samples of tetracopper calcium oxychloride, precipitates from slowly hydrolized ferric chloride solutions, and tungsten oxide to illustrate some of the effects which may occur. The polymorphism of these substances, particularly of tungsten oxide, is discussed.

The major effects which contribute to the pseudostructure formation are melting, evaporation, change of crystalline states, increase in degree of crystallinity, and migration of material. Reasons are given for concluding that the primary cause of specimen changes is to be found in the heating effect of the beam. The possibilities of secondary causes are also considered.

Artifacts may also be introduced into bacterial specimens. Observed changes in such samples are described briefly and examples are given.

IT has been apparent for some time, particularly since the advent of the biased gun, that electron microscope specimens are often affected in many ways by electron bombardment.[1] As long as the biased electron source* was not widely used, the occurrence of pseudostructures was not of great importance, but now that the majority of laboratories are turning to them, a discussion of their effects upon specimens is in order. This paper deals with the particular phenomenon of the pseudostructures which may be introduced into the object under the beam and, hence, into the micrographs. Several examples are given in order to illustrate the artifacts which can be induced and to emphasize the necessity for care when examining specimens and interpreting their images. The term *pseudostructures* is defined here as those structures which are not real properties of the samples but are artifacts introduced by the electron bombardment.

INORGANIC CRYSTALS

In Fig. 1 specimen changes are illustrated in a dry, powdered sample of *tetracopper calcium oxychloride*** which is supported upon a Formvar film. This material is affected by electron beams of even very low intensity which will introduce slight granulations into the platelets. These granulations are observed in the platelet at the arrow in Fig. 1A in spite of precautions which were taken. When the intensity is raised, the startling changes seen in Fig. 1B occur. With extreme rapidity, the platelet particles and aggregates become groups of round particles when a sufficiently high illumination is reached. It will be observed in Fig. 1B that the outline of some of the original particles is left in the substrate and that a particle dispersion directed away from the affected particles has formed over the film.

In a dispersion of tetracopper calcium oxychloride, large and small groups seem to be affected equally. The changes will occur at relatively low intensities but they take place with even greater rapidity at higher illumination. From observations it is concluded that a combination process of melting and vaporization takes place, and that the vaporized material is redeposited upon cooler portions of the film to give the effect of shadow casting in Fig. 1B. The appearance of the unbombarded material in this sample is always such as to suggest that it is crystalline, but Bragg reflections are noticeably absent from any but the bombarded areas. This indicates that a change in crystalline habit has occurred under the beam and that lattice orientations have been introduced which are suitable for producing numerous, strong reflections.

In order to test the effect of shadow casting

[1] E. F. Burton, R. S. Sennett, and S. G. Ellis, Nature 160, 565 (1947).

* Available from Vibration Systems, Inc., 1040 W. Fort Street, Detroit, Michigan.

** Sample received from E. I. DuPont de Nemours and Company, Wilmington, Delaware.

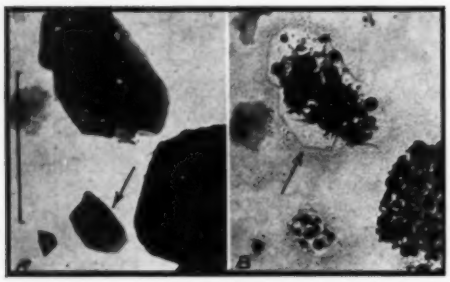


FIG. 1. Tetracopper calcium oxychloride, ×20,000. (A) At low intensity. Note granulation already appearing in the platelet particle at the arrow. (B) Same field at high intensity after an extremely short bombardment. Note at arrow the outline of the original particle left on the film and the dispersion of very fine particles.

upon the formation of pseudostructures, a sample of tetracopper calcium oxychloride which was mounted in a manner identical with that of the specimen in Fig. 1 was shadowed with chromium. It is found that the extent of the changes and the speed with which they take place are diminished, and that no effect at all can be demonstrated in most of the bombarded regions of these shadowed samples. Although illuminated at high intensities for considerable periods, even the large aggregates are relatively much less affected. However, it is possible finally to cause the metal itself to migrate under intense illumination.

FIG. 2. The precipitate from a slowly hydrolized solution of ferric chloride, ×20,000. (A) At low intensity. (B) A different field at higher intensity, showing granulation in the particles and over the film and a track left by a migrating melted particle. (C) Another field at even higher intensity in which most of the particles have become rounded units after melting.

The explanation for the diminished effect in shadow-cast specimens probably lies in the fact that the metal coating is able to conduct away a large part of the heat which otherwise would be built up. In addition, the metal tends to prevent migration and evaporation of the material and serves to protect it to some extent from other effects which, theoretically, the beam could have upon it.

In Fig. 2A a specimen is shown in the unchanged condition of the *iron oxide* precipitate formed by a slow hydrolysis for three months at room temperature of a 3/10 percent solution of ferric chloride.[2] This specimen was mounted by drying a small drop of the suspension on Formvar. The material exists in an unstable crystalline state, and in its condition here the particles are composed of bundles of platelets lying face-to-face. Figure 2B shows the appearance of the material after a partial change has taken place at a relatively low intensity. Gross granulation has occurred, and some melting and evaporation have begun. When the particles melt, they are observed in the electron microscope to migrate as rounded units, leaving tracks over the supporting Formvar. The Formvar film does not break. Such tracks are exampled at the arrows in Figs. 2B and 2C. The evaporated substance has redeposited on the substrate to give an appearance similar to a light shadow-cast of the film. In Fig. 2C a more intense bombardment has left only rounded units behind. There is evidence of tracks in the film and granulation over the surface due to the deposition of evaporated material.

Figures 3A, B, C, and D show the appearance of a particular field of a *tungsten oxide**** sample in a series of electron micrographs taken after various amounts of intense electron bombardment. In order to obtain this series of plates, the illumination was raised until a change began to take place, then a short period of bom-

[2] H. Zocher and W. Heller, Zeits. f. anorg. u. allgew. Chemie **186**, 75 (1930).

*** The details related to the chemical preparation [H. Zocher and K. Jacobsohn, Kolloid Beih. **28**, 169 (1929)] of the samples will be given in a forthcoming publication in the field of tactoid chemistry.

bardment was allowed before the intensity was again lowered and a micrograph taken. Figure 3A shows two flat, structureless platelets of the oxide superimposed upon a background of unresolved, particulate material. This micrograph was taken at very low intensity and no known changes have occurred. Under bombardment in Fig. 3B, a peculiar migration of the background material takes place although it remains particulate, and the platelets develop a gross structure. In Figs. 3C and D the original platelets are evaporated almost completely away and the background becomes a lace-like material made up of very tiny crystals. The supporting film is destroyed finally, and the appearance of the lace-like areas is illustrated at higher magnification in Fig. 4.

Whereas in the other two samples the effect of bombardment was to change large particles into smaller units, in this sample, while the large particles are being converted into smaller structures the particulate background material is experiencing a growth phenomenon. It would appear, therefore, that either process can be initiated by electron bombardment.

Figure 5A shows a bombarded group of the tungsten oxide plates where the structures are more easily visible at higher magnification. An interesting mosaic pattern of rods is observed. In dark field[3] these rods are seen to be highly crystallized, and Bragg reflections are very predominant. Figure 5B, in focused dark field, shows the same area as Fig. 5A. In unfocused dark field, Fig. 5C, the Bragg reflections are made more obvious.

These rods are rendered visible under bombardment by a process of selective evaporation over the platelet, or migration and recrystallization, or by a combination of both mechanisms. The final explanation is probably closest to the truth since the bombardment not only produces Bragg reflections and hence presumably initiates a new crystalline state, but if continued long enough, it causes the whole platelet to evaporate.

It is known from other electron microscope observations of these tungsten oxide samples that they are unstable in crystalline habit. For example, if a suspension of the platelets is left

[3] Cecil E. Hall, J. App. Phys. 19, 198 (1948).

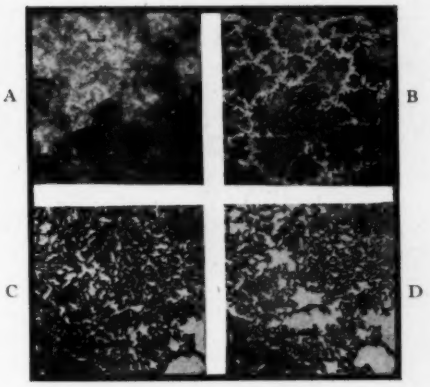


FIG. 3. Tungsten oxide, ×2000. This series was taken after increasingly more exposure to an intense electron beam. The original somatoid first develops a structure and finally evaporates; the background particles migrate and form larger units and crystals.

for a certain time and examined at intervals, it is found that their appearance changes spontaneously from that in Fig. 6A to that of Fig. 6B. Electron bombardment apparently hastens this process of change and further illustrates the polymorphism of the compound. The same effect which is visible in Fig. 6 can be initiated more rapidly by simply washing and centrifuging the suspension a number of times in water.

The platelet particles of Figs. 3A and 6A have been known as *somatoids*,[4] and from these electron microscope studies it is concluded that they are polymorphic, approaching the more stable rhombohedral form on standing in suspension or being washed in water. The crystals which are

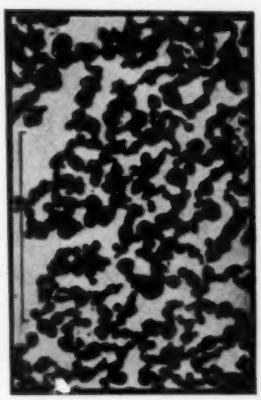


FIG. 4. A section of Fig. 3D at a higher magnification, ×25,000.

[4] V. Kohlschuetter, *Die Erscheinungsformen der Materie* (B. G. Teubner, Leipzig, 1917).

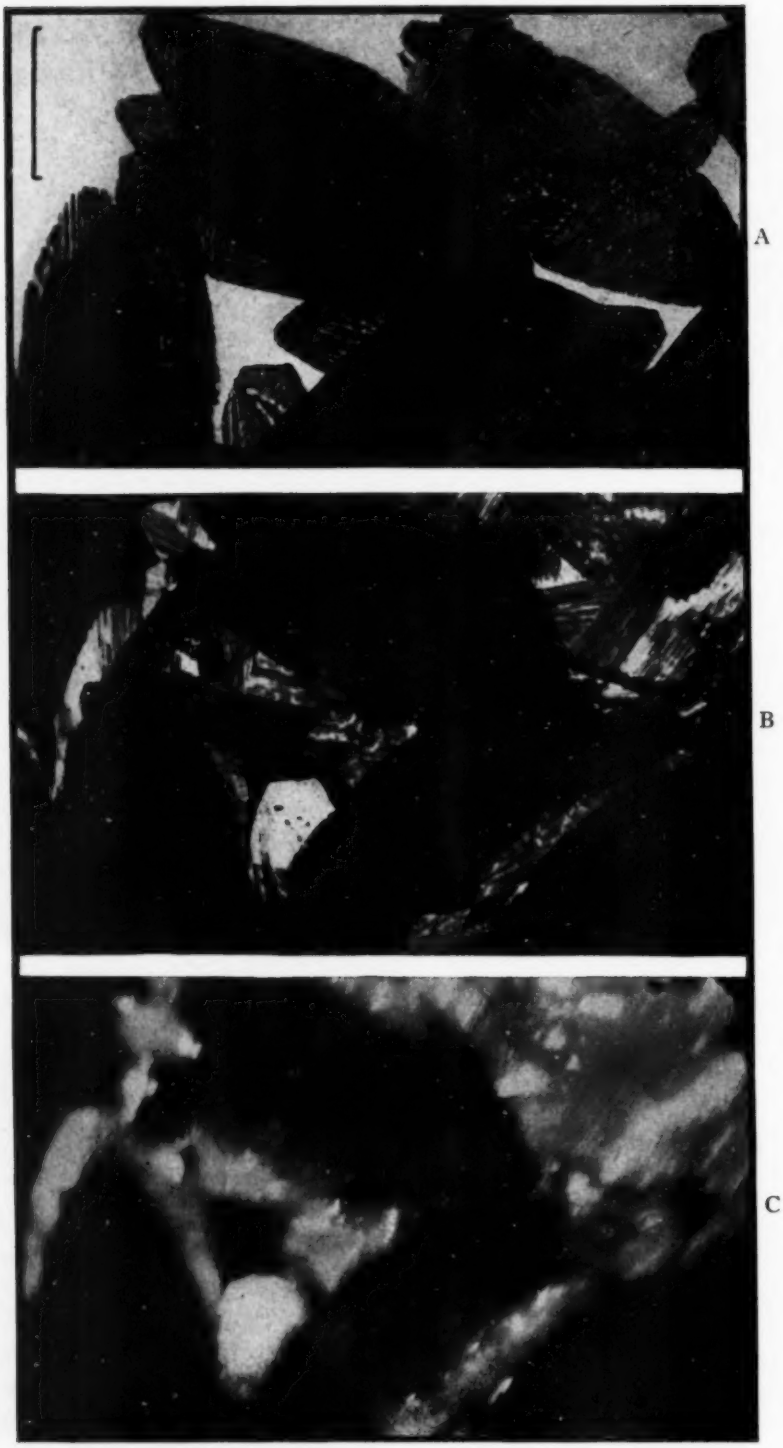


FIG. 5. Tungsten oxide somatoids after electron bombardment, ×20,000. (A) Bright-field electron micrograph showing the mosaic rod pattern formed in the somatoids. (B) Dark field of the approximately same group in focus. (C) Same group in dark field out of focus to show Bragg reflections.

formed spontaneously on standing are much thinner than the original somatoids and are straight-edged where the somatoids have rounded edges. The somatoids are affected by electron bombardment, as has been described, but no effect except for slow evaporation is noted in either the rods in bombarded somatoids or the thin crystals which form on standing.

The tips of the original somatoids are frequently observed to have toothed edges. The height of these steps averages about 380A. On the other hand, the thickness of the rods formed on bombardment averages about 360A. This close agreement in size, plus the fact that the orientation of the steps before bombardment is the same as that of the rods seen in the same particles after bombardment, indicate that the pattern of the mosaic is dictated by the manner in which material is laid down in the original somatoid.

A number of further general observations concerning pseudostructures can be made before explanations are offered for the observed phenomena. It is often noted that the changing of one area seems to act as a "trigger" for the setting off of the change in neighboring areas. In some specimens it is difficult to initiate the production of pseudostructures at all, but when they are introduced into one locality nearby regions quickly respond.

The changes do not always happen instantaneously but may develop progressively with time or intensity. Some specimens require a focused beam and a fair length of exposure before changes occur, but more often they take place with startling rapidity at a limiting intensity somewhat less than that at focused beam. Thus, although the initial introduction of structure into the tungsten oxide particles takes place extremely rapidly at high intensity, it can be introduced more slowly at lower intensities. At the same time the gradual evaporation is a progressive thing over a much longer period. Therefore, while a biased gun seems to be the chief offender, it is not unreasonable to suppose that to some extent the same effects can occur when an unbiased one is used.

Electron diffraction patterns taken before and after bombardment indicate in those substances which are changed by the beam either an increase in crystallinity or a change in crystal structure, or both effects simultaneously. These observations are supported, in the case of the iron oxide precipitate, by the earlier reports[5, 6] that these materials change from β-FeOOH to α-FeOOH spontaneously on standing, or by variations in experimental conditions, and notably under conditions of *heating*. Marked improvement and change in the crystalline habit is also indicated by the prolific production of Bragg reflections in bombarded regions. Thus, the detection of crystalline occurrences by observations of Bragg reflections may also be complicated by electron bombardment, and pseudo-observations may be made of these characteristics in electron images.

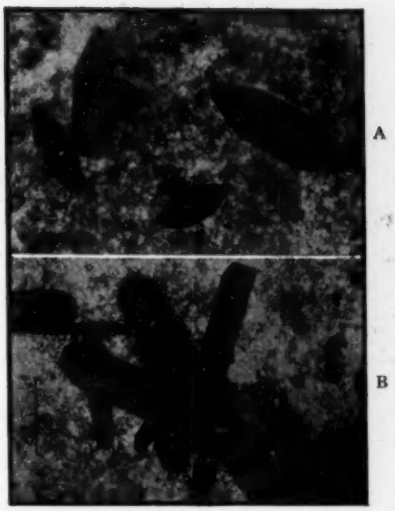


FIG. 6. Tungsten oxide unbombarded, ×10,000. (A) Somatoids from a freshly prepared sample. (B) Orthorhombic crystals formed spontaneously in same sample after standing 23 days.

Figure 7 is included as a further example of the dark-field appearance of bombarded speci-

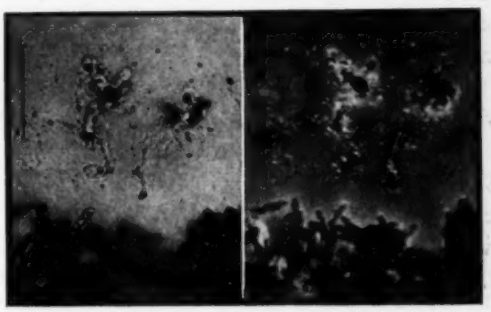


FIG. 7. The precipitate from slowly hydrolized ferric chloride solution after bombardment, ×12,500. (A) Bright field showing apparently amorphous track residues and particles. (B) Same view in dark field showing Bragg reflections from this bombarded area.

[5] W. Heller, O. Kratky, and H. Nowotny, Comptes Rendus **202**, 1171 (1936).
[6] O. Kratky, and H. Nowotny, Zeits. f. Krist. **A100**, 356 (1938).

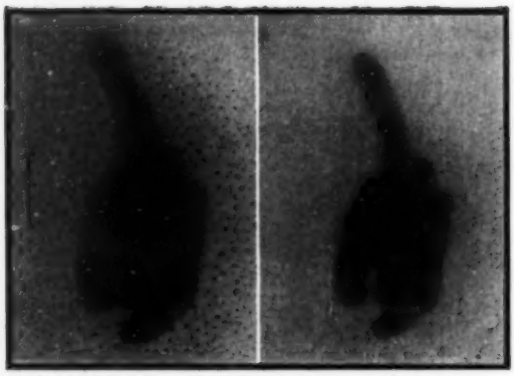


FIG. 8. Bacillus *tuberculosis*, ×10,000: (A) at low intensity; (B) same field after intense bombardment.

mens. Figure 7A is a bright-field micrograph of a bombarded sample of the precipitate of iron oxide from iron chloride. In the dark-field micrograph of the same view, in Fig. 7B, the electron reflections are seen coming predominantly from the melted and redeposited material, although even the partially changed groups near the base of the micrograph are causing some reflection. In samples of this material examined at low intensity with no apparent change from their normal state, no Bragg reflections are recorded at all. The same observations are made concerning the tungsten oxide and the tetracopper calcium oxychloride specimens.

The fact that such refractory materials as titanium dioxide are melted under the beam indicates that the heat developed in *local* areas is sufficient to cause specimen changes. Large chunks of titanium dioxide with indefinite shape are changed to round units which run together, which may be distorted if disturbed and which apparently remain molten as long as they are illuminated at high intensity. They may be melted again if the beam is redirected upon them. Small particles of titanium dioxide are completely unaffected. Since the melting point of SiO_2 is 1640°C, it appears that local temperatures of at least the order of 1600°C can be achieved in electron microscopes, using a biased gun. On the other hand, it is interesting to note that carbon black particles are not observed to sublime or develop pseudostructures under the beam during periods of normal observation, and no visible evidence of graphitization is observable in bombarded carbon black particles at any intensity. Graphitization temperatures are in excess of 2000°C, and usually reach 2500 to 3000°C.

Although the Formvar films would be destroyed if a large part of one field were raised to such a temperature, such breaks are not observed unless the particles are very large when the film is broken, probably more by movement of the particle than by any direct effect of the beam. It should be emphasized that the temperature rise is very localized, existing only *at* the particle which is being bombarded, and it depends upon the amount of energy this particle is able to absorb.

It is also possible that activation or ionization by the beam could play an important role in the introduction of the changes, to lower melting points, to initiate chemical reactions, to cause crystal changes or ionic migrations. However, if these phenomena do occur, since the individual electron energies are the same from both the biased and the unbiased gun, it is difficult to explain why at least somewhat comparable effects are not observed with either an unbiased gun at high intensity or a biased gun at low intensity.

From the observations it appears that heat

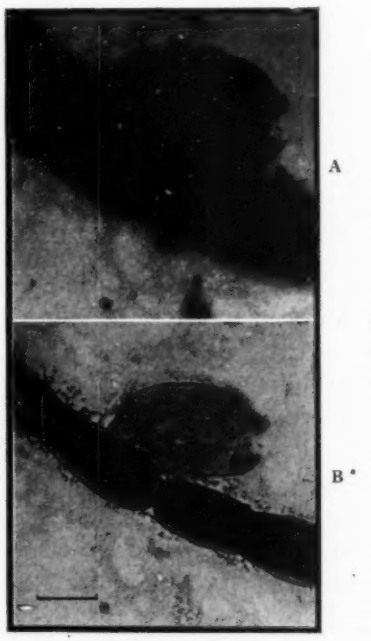


FIG. 9. Bacillus *subtilis*, ×8000: (A) at low intensity; (B) same field after intense bombardment.

generated by the electron beam is the *prime* fundamental cause of the phenomena of pseudostructures. The reasons for making this statement are the following: (1) in all samples where pseudostructures are noted or where changes of shape and morphology occur, the effects are much enhanced as the intensity is raised, (2) the larger and thicker particles of a dispersion which would be expected to be able to absorb the most heat are indeed changed most quickly and efficiently by the bombardment, (3) the effects are more often observed using a biased gun from which the beam is more concentrated, (4) many substances are melted, (5) there is production of crystalline structures, and (6) no effect usually occurs until a certain limiting gun intensity is reached. These reasons are stated generally, with the knowledge that the importance of any one of them varies considerably with the properties of the specimen itself.

Since the effects often occur rapidly and almost instantaneously after a limiting intensity is reached, it may be that the beam is able to introduce effects other than heating which tend to catalyze the changes in some specimens. However, this again seems to be a property dependent to a large extent upon the specimen itself since some materials do not change at all and others (e.g., flowers of sulfur or tungsten oxide in the stable, rhombohedral form) merely sublime, regardless of intensity or length of bombardment. Undoubtedly, the inherent properties of the specimen govern the changes which occur, and substances which are unstable crystals, or are polymorphic, seem to change more rapidly and easily than others.

From the observations recorded here, it appears that the development of pseudostructures can have indirect advantages in electron microscopy for (1) the study of polymorphic forms, (2) investigation of the construction mechanisms in certain unstable crystal forms, and (3) detecting the presence of such unstable states.

On the other hand, in normal electron microscopy where artifacts are not desirable, the procedures of examination should be such that original bombardment of all samples is very light. The object should be micrographed under these conditions and if the illumination is raised, careful observations should be made of any specimen variations which take place. These are perhaps the major criticisms of the biased gun: that greater care has to be exercised to ensure against specimen deterioration or change and that the intensity from it is uncontrolled except by means of condenser current. However, we have found no specimen which could not be micrographed successfully with it.

BIOLOGICAL SPECIMENS

Since the biased gun is coming into such universal use, its effect on certain biological specimens was studied. Some of the more important observations and conclusions made in this investigation are given briefly in the remainder of this paper.

Pseudostructures are observed in organisms which are examined under electron bombardment. The effects may take many forms but any one or all of the following artifacts may be produced: (1) enhancement of structure already dimly visible over the body of an organism, (2) introduction of what may appear to be a capsule where none was visible before bombardment at lower intensity, (3) formation of round distinct vacuoles within the granules of an organism, (4) changes in the diameters of the granules, (5) shrinkage of cells in both length and width, (6) artifacts introduced by deposition over the organism of material evaporated from neighboring debris, (7) shrinkage of cell content so that it pulls away from the wall, (8) in some cases complete disappearance of all cell content, and (9) specimen charring. Sometimes small distinct vacuoles characteristic of the type formed in granules will become visible in regions where no granule was previously observable. The formation of such vacuoles may be an indicator of the location of materials which are concentrating in the process of forming a granule.

Two examples of the effects of electron bombardment upon bacteriological specimens are given in Figs. 8 and 9. Figure 8A is a field of unbombarded *Mycobacterium tuberculosis* in which granules may be plainly seen. Figure 8B shows the same field after bombardment at higher intensity for about thirty seconds. The background material has evaporated and one or more vacuoles have formed in the granules. These are usually produced eccentrically, and as many as

three have been observed in the same granule. Almost invariably the cells are found to shrink about ten percent under bombardment in vacuum.

These vacuoles make their first appearance suddenly at a limiting intensity, but the major part of their growth is gradual over the period. When most of a granule has disappeared, continued bombardment produces no further effect upon it. This indicates that when the mass of a granule has been dissipated sufficiently the amount of energy absorbed is no longer able to yield temperatures high enough to do damage, and, consequently, that it is the heating of the organism again which is causing the observed changes.

Bacillus subtilis is illustrated in Fig. 9A before any change has taken place and in 9B after bombardment. In this sample the cells shrink, the debris evaporates, cell content pulls away from the wall, vacuoles form and, where there is no visible capsule before, there is one after the bombardment.

The methods by which bacterial specimens are mounted will affect the pseudostructures materially. Wet samples examined soon after preparation are affected most, while those mounted dry or left in air are affected least. It is often impossible to introduce visible artifacts into preparations which have stood in air for some days, although the same materials when fresh are easily distorted in one way or another. Regardless of mounting procedure, there is a tendency for cells to shrink when illuminated at high intensity in a vacuum. This is an important point to be kept in mind when doing size distribution studies. These facts indicate that the observed changes listed above are due chiefly to evaporation of contained liquid which is able to diffuse through the wall without causing rupture.

The freer a preparation is from debris, the less are the chances for the introduction of artifacts. Debris in well dried samples is affected less than that in fresh or undried samples, but certain materials, such as sodium chloride, will always evaporate under the beam and distort the object. The cells from young cultures seem to be more affected than those from older growths, but this point is not yet clearly established.

It is clear that when one is interpreting morphological characteristics of biological specimens from electron micrographs the effect of electron bombardment on the samples is another factor to be reckoned with.

ACKNOWLEDGMENT

The author is indebted to the Department of Laboratories of the Henry Ford Hospital for the use of their electron microscope for these investigations. Bacterial samples were prepared by Miss Edna Jackson of the Department of Laboratories. Samples of tungstic and iron oxide were studied in collaboration with Dr. W. Heller and Mr. W. Wojtowicz of the Chemistry Department, Wayne University.

A Range-Finder Testing Laboratory

IRVINE C. GARDNER
National Bureau of Standards, Washington 25, D. C.
(Received January 23, 1948)

A laboratory for testing optical range finders constructed under the sponsorship of the Ordnance Department, U. S. Army is described. The central piece of the equipment, specially designed and constructed for the purpose, is a testing instrument that optically simulates targets at ranges from 855 yards to infinity, introducing the corresponding angular parallax between two beams of light at a separation of 15 feet, with angular errors not exceeding one second. This instrument is installed in a room with temperature control equipment and also provision for applying directed radiation to simulate extremes of temperature substantially as great as those encountered under service conditions. Such a laboratory provides convenient means of rapidly acquiring exact information regarding the behavior of a range finder and localization of the sources of error. It also enables one to predict how a given range finder will perform at any point on the earth's surface without the necessity of maintaining laboratories in the different temperature zones.

I. INTRODUCTION

A RANGE finder[1] is perhaps the most exacting and complicated optical instrument that is built, and for its successful operation all parts, optical and mechanical, must be designed to eliminate error or introduce compensating characteristics. A successful range finder must measure angles with errors not exceeding one or two seconds. It must do this despite its large size (from 3 to 50 feet long) and weight, despite the rough treatment that it necessarily receives when transported from place to place under war conditions, and despite the fact that measurements must be made under conditions that make it inconvenient or impossible to check the zero or average a series of observations to obtain a better value. Consequently, a range finder is necessarily so designed that the primary errors resulting from flexure or expansion of optical and mechanical parts and from the temperature-induced changes of index of refraction of the gas filling the interior of the instrument are eliminated or compensated. As a result of careful design the residual errors are small and may result from such subtle causes as heterogeneities of material which cause differential expansion, temperature gradients produced by radiation or lack of provision for symmetrical heat flow to or from critical parts, changes in flexure produced when the orientation of the instrument with respect to the gravitational field is changed, and temperature gradients in uncompensated columns of gas through which the light passes within the instrument. The number of sources of error of this nature is so large and the individual errors are so small and so nearly equal in magnitude that it is quite impossible to separate and identify them, unless observations can be made with adequate control of all ambient conditions such as temperature, exposure to radiant heat, and so forth.

Consequently, to make a detailed analysis of a range finder, which will give accurate information concerning the sources of the different sorts of errors and their magnitude, is quite impossible

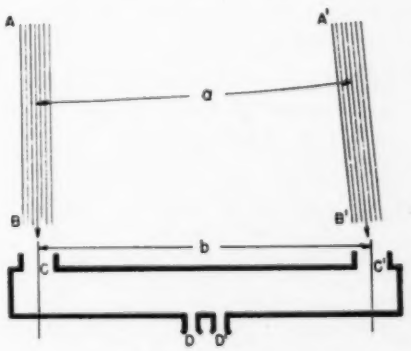


FIG. 1. Diagrammatic sketch showing functional operation of range finder.

[1] The term *optical telemeter* applies to a range or height finder. A height finder is a range finder with a computing device, either mechanical or optical, to multiply the slant range by the sine of the angle of elevation to yield directly the height of the target. The height finder therefore has, in general, all the sources of error of the range finder plus those introduced by the additional computing device. The height finder is obsolescent in the army, and the term optical telemeter has never come into general use in this country. Therefore, throughout this discussion the term range finder will be used with the general understanding that any statement also applies to height finders.

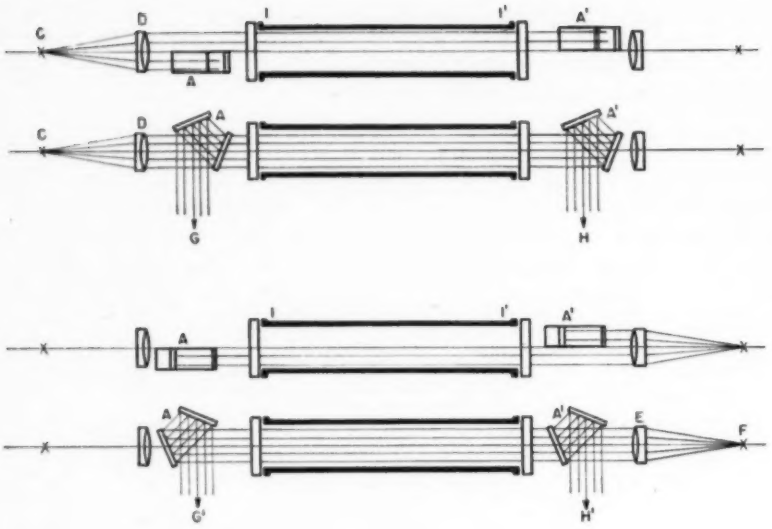


FIG. 2. Diagrammatic sketch showing principle of operation of testing instrument. Upper pair: elevation and plan showing optical squares in position for operation of left collimator. Lower pair: elevation and plan showing optical squares in position for operation of right collimator.

by means of an observation program of practicable length in which terrestrial targets are used. Furthermore, fixed terrestrial targets are seldom available at altitudes greater than 5° or 10°, and experience has clearly shown that tests at low elevations are unsatisfactory criteria for the performance of the range finder on targets at greater angular elevations. The use of airplanes suggests itself as a suitable means of providing elevated targets, but the true altitude and range are required in tests and therefore airplanes are not satisfactory, unless a costly installation of photo-theodolites manned by a crew of observers and computers is available for all periods of test. Moreover, even if the difficulties incident to the close control of ambient conditions and the provision of suitable targets at all elevations were surmounted, natural targets would still be disadvantageous because the long column of air between a range finder and a distant target, as a result of its heterogeneities, introduces an imprecision of measurement that makes many more observations necessary than would be required if a laboratory target were used.

In view of these conditions, and in order to meet the requirements of the range finder designer, a complete unit for testing range finders, sponsored by the Ordnance Department of the Army, has been established at the National Bureau of Standards. The central feature of the laboratory is a test chamber which can be maintained at any temperature within the range 0° to 100°F, or in which a given rate of temperature change can be maintained. In this test chamber there is the precision collimator testing instrument, which may be described, functionally, as a collimator system that produces a simulated target suitable for observation with a range finder. Means are provided for presenting the target to the range finder at several different known ranges and at any elevation from 0° to 90°.

To make clear the performance required of an optical device that simulates a distant target for the purposes of range finder testing, a brief description of the functional performance of a range finder will be given. The range finder is essentially a double-telescope system with the entrance pupils of the two telescopes at C and C' in Fig. 1 and with the oculars near the center of the instrument at D and D'. Two beams of light entering the range finder are indicated at AB and $A'B'$. These beams of light, which are not parallel but include the angle α, are to be considered as proceeding from a target at a range, r, so great that the target cannot be indicated on the diagram. If b is the separation of the entrance pupils, it is evident that

$$\alpha = b/r. \tag{1}$$

The length b is termed the base length of the instrument. The determination of a suitable base length for an instrument to serve a given purpose is usually determined as a compromise. As the

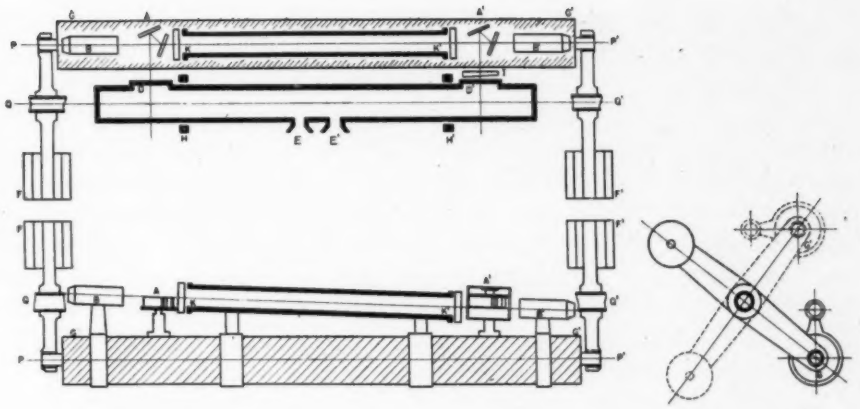


FIG. 3. Diagrammatic sketch of range finder testing instrument. Upper: Plan with range finder in position. Lower: Elevation with range finder not in position. Right: Positions shown for testing at 0° (full line) and 90° (dotted line) elevations.

base length is increased the range error corresponding to a given error in the measurement of α decreases and, consequently, the precision with which the range is measured increases. On the other hand, as the base length is increased the portability of the instrument is decreased. Usually the compromise is such that the value of α for the greatest and least ranges represented on the scale are approximately 10 seconds and 30 minutes, respectively. This corresponds to maximum and minimum ranges of approximately 20,000 and 100 base lengths. Unfortunately, the longer ranges corresponding to the smaller values of α are the more important, and it is for this reason that the range finder must necessarily measure angles with the greatest possible accuracy.

II. FUNCTIONAL REQUIREMENTS AND PRINCIPLE OF DESIGN OF TESTING INSTRUMENTS

From the foregoing it is evident that an apparatus for producing a simulated target suitable for a range finder must provide two beams of collimated light separated by the base length of the instrument to be tested (a maximum of 15 feet for the testing instrument that has been built) that shall be parallel or depart from parallelism at will by known values between zero and thirty minutes, the exact angle to be known in all instances with an error certainly not greater than one second and preferably less. Furthermore, in order to simulate elevated targets, it must be possible to direct the parallel or nearly parallel beams of light horizontally or downward at any one of a selected series of angles ranging from 0 to 90°, this change in apparent elevation being effected without altering the included angle between the two beams.

The design of an instrument to accomplish this purpose as outlined is made particularly difficult because of the great distance between the two beams of light. It is quite impracticable to make a structural beam of the required length sufficiently rigid to maintain the angular relations with the required accuracy and of weight such that it can be moved to simulate a target at different elevations. Accordingly the point of departure for the design of such an instrument is the absolute collimator system with two collimators originally used in Zeiss range finders for making the so-called infinity adjustment. This system is shown diagrammatically in Fig. 2. The upper pair of drawings shows an elevation and plan of the optical system. The lower pair of drawings is the same except that parts A and A' have been rotated, as will be described later.

Considering the upper pair of drawings, the reticle C and the objective D form a collimating system which provides a beam of parallel light. At A and A' are two optical squares, each consisting of two accurately plane front-surface mirrors including an angle of 45°. In the plan view the two reflectors are placed, one below the plane of the drawing, the other above it so that each intercepts half of the beam from the collimator. This is shown in the elevations. Nominally each optical square deviates half of the original beam through 90° and the directions of the two beams, as indicated by the arrows at G and H, are parallel. In general, because of variation of workmanship, the directions of the beams at G and H are not parallel with the pre-

cision required for height finder testing. In the lower pair of diagrams of Fig. 2 the collimator comprising the objective E and the reticle F is utilized, and the optical squares A and A' are rotated through approximately 90° as indicated. The two beams of parallel light now emerge in the directions G' and H'. If the deviations produced by the optical squares at A and A' remain constant for the two arrangements of Fig. 2, the beams G and H for the first position and G' and H' for the second position deviate equally but in opposite senses from parallelism. Consequently, if a range finder is used on the targets presented by the two positions of the optical squares, the average of the two readings is the value which would be obtained if directions G and H or G' and H' were rigorously parallel. The targets in the two collimators are cross-hairs, opaque on a bright ground, with each "hair" inclined 45° with respect to the vertical in the field of the range finder. This type of target is favorable for precise setting with the vertical markings found in most stereoscopic range finders. At I, I' there is an evacuated tube closed by the plane parallel end windows. If this tube were not present, transverse temperature gradients in the air between the two optical squares might cause the beams of light to follow a curved path, thereby introducing a systematic error.

III. DETAILED DESCRIPTION OF TESTING INSTRUMENT

Figure 3 shows the general arrangement of parts diagrammatically. The upper view is a plan view when the parts are in position to simulate a target at zero elevation. The range finder is shown with entrance pupils at D and D' and oculars at E and E'. The lower view of Fig. 3 is an elevation, the range finder not being represented because it would obscure the essential optical parts of the testing instrument. The optical squares A and A' and the two collimators B and B' are carried on the large tube GG' which is the rigid support for the entire optical system. The optical squares and collimators are attached to the main tube by clamps which permit them to be adjusted along the length of the tube to adapt the instrument to range finders of different base lengths. As can be seen in the elevation, the two optical squares are mounted at the same distance from the tube GG'. As a consequence, the optical squares are at the same height from the floor and can be simultaneously centered with respect to the two entrance pupils of the range finder. Because of this manner of mounting the optical squares, the line of centers of the collimators is necessarily inclined to pass over one optical square and under the other. The axes about which the optical squares rotate are normal to this inclined light path. Tube K, K', which is between the two optical squares and also inclined from the horizontal, is evacuated. As has been noted, this prevents systematic errors caused by stratification of the air. The vacuum tube also contributes materially to the steadiness of the image by eliminating the effect of "boiling" when the temperatures of the range finder and the ambient air differ greatly.

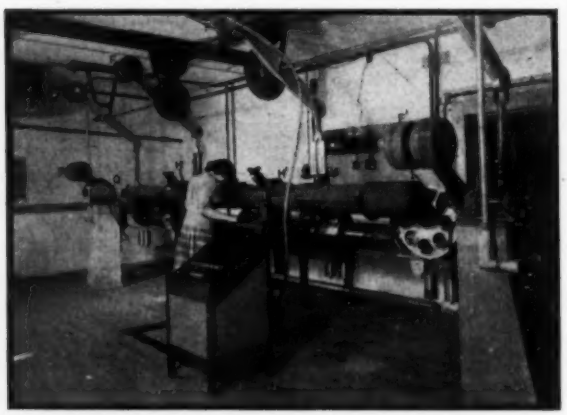


FIG. 4. Testing instrument with $M1$ height finder in position for test at 0° elevation.

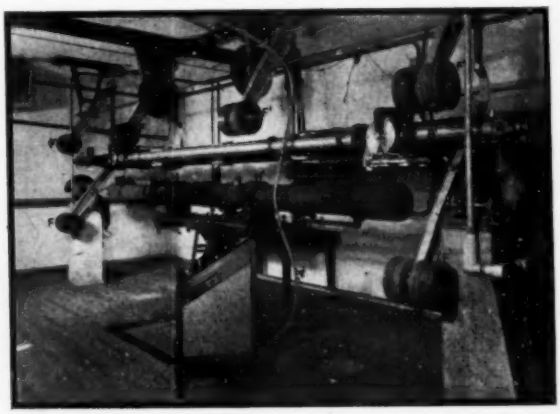


FIG. 5. Testing instrument with $M1$ height finder in position for test at 90° elevation.

The tube GG', which is shaded in Fig. 3, is the main structural member of the system. It is a seamless drawn steel tube 8 inches in diameter, with a wall-thickness one-eighth of an inch. In order that this tube may be free from constraint it is supported by two coaxial spindles P and P', which in turn ride in self-aligning bearings. These bearings are in the ends of two lever arms which rotate, respectively, about the trunnions Q and Q'. The weight of the optical system and supporting tube is counterbalanced by the weights F and F' at the ends of the lever arms. The height finder to be tested, with entrance pupils at D and D', rotates about its longitudinal axis in the bearing rings H and H', as for viewing elevated targets. In the position for test the height finder is aligned with its longitudinal axis of rotation coincident with the axis of the trunnions Q and Q'. In order that this adjustment may be made with nicety, the range-finder tripod rests on a large iron plate, flush with the floor, which can be shifted horizontally on ball bearings by two slow-movement devices acting at right angles to each other. The adjustment in height can be obtained by the leveling screws of the tripod. By rotating the testing instrument about the trunnions Q and Q', a target at any elevation can be simulated and the targets can be followed by the height finder by rotating it in the bearing rings H and H'. To simulate a target at a finite range, an optical wedge producing the required angular parallax is introduced into the beam of light proceeding from one of the optical squares to the height finder. Such a wedge is shown at I. The maximum deviation to be produced by a wedge does not exceed 30 minutes (corresponding to a simulated range of 100 base lengths of the instrument). Because of the smallness of this angle it is not difficult to control the positioning of the wedge with sufficient invariance to maintain constancy of the simulated range with the required precision.

The instrument as actually constructed is shown in Figs. 4 and 5 on which the parts corresponding to those shown on Fig. 3 are identified by the same letters. Figure 4 shows the position of the instrument for simulating a target at 0° elevation, the optical parts, in this instance, being largely concealed by the $M1$ height finder which is shown in position for test. Figure 5 shows the position for a target at 90° elevation. To secure the required accuracy, it is necessary to incorporate many details of construction that have not been mentioned in the discussion of Fig. 3. The weight of the 8-inch tube and optical system is so great that it did not appear practicable to carry the counterbalancing weights on a single pair of levers. Consequently, four overhead levers are connected to the two lower ones by linkages so that all remain parallel for any elevation for which the instrument is set. There are, therefore, six groups of counterweights, four on the overhead levers and two on the lower levers. Each group is denoted by F or F' in Figs. 4 and 5. The four upper levers for counterbalancing are not mounted on a single shaft but on two lengths connected by large spur gears and a jackshaft. This combination is utilized in order to avoid mechanical interference with the overhead monorail which carries a differential hoist by which height finders are installed in or removed from the testing apparatus.

The 8-inch tube which serves as a beam to support the entire optical system of the instru-

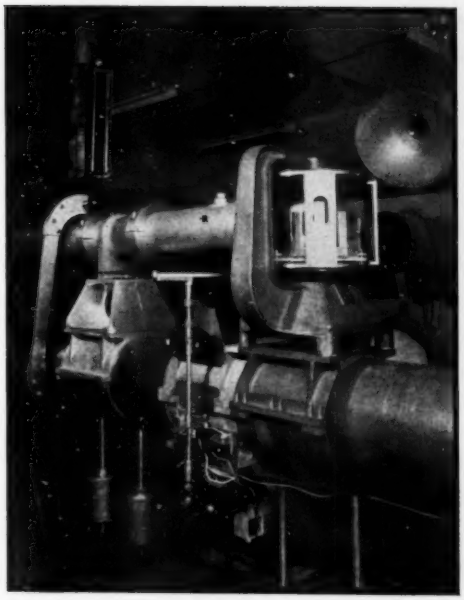


FIG. 6. Rotating table support for one of the optical squares. The electric drive for the rotating table and a portion of the vacuum tube extending between the two optical squares can be seen. At the left are the three spring balances and the large hook which form one of the intermediate supports of the 8-inch main tube shown in the lower half of the photograph.

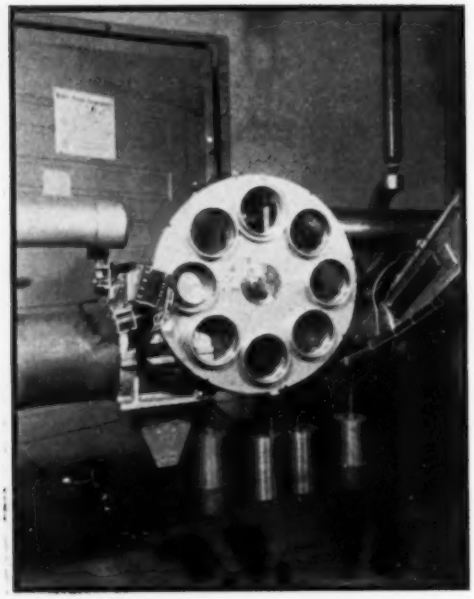


FIG. 7. The turret by which differing angular parallaxes are introduced to simulate different ranges. The large disk contains eight apertures, of which one is clear and the others contain achromatic optical wedges by which the beam of light entering the right end of the range finder is deviated. At the left can be seen the limit switches which control the driving motor and stop the disk in a position corresponding to any one of the notches in the periphery of the disk.

ment sags under its own weight causing the axes about which the two optical squares rotate to lean toward each other. If this flexure is not compensated or reduced, the path of a ray of light through the optical square will not lie in a principal plane and the deviation will not have the correct value. This cannot be corrected by readjusting the axes of the optical squares, because when the instrument is set for zero elevation the full component of the flexure contributes to the displacement of the axes of the optical squares, but as the elevation of the target is increased this contribution becomes less and vanishes completely for 90° elevation. The flexure of the beam is reduced to a satisfactorily low value by introducing two intermediate supports to the beam, one of which extends down from each of the two inner overhead counterbalance levers. A spring dynamometer made from three heavy spring balances combined in parallel is inserted in each of these intermediate supports and the tension carefully adjusted in order that the upward force applied at each point of support may carry the weight of the tube to such an extent that the parallelism of the axes about which the optical squares rotate may be substantially independent of the particular elevation for which the testing instrument is set.

In the discussions of the diagram of Fig. 2 it has been mentioned that the precise working of the instrument involves the assumption that the deviations introduced by each optical square remain constant during the period required for making the observations in the two positions of Fig. 2. This is not a very severe requirement, because it takes less than a minute to take the required pair of observations. However, it is desirable to have the optical squares as invariant as possible with respect to temperature, as this enables the readings taken over a period of time to be checked more readily and also because it is convenient to operate without the very frequent reversals of the optical squares that are required if their stability is suspected. Furthermore, the reflecting surfaces of the optical squares must remain plane during the progress of the test at different temperatures, as otherwise a large amount of parallax will be introduced between the reticles of the height finder under test and the simulated target. To satisfy these requirements, the optical squares, which are of the mirror type, are made from three pieces of fused quartz, two mirrors and a base, which are united by optical contact. The mirror surfaces are coated with evaporated aluminum. Separate tests on optical squares made in this way have shown that the performance is entirely satisfactory for the use which is made of them.

Making observations with a range finder is an exacting task if precise readings are to be obtained. It is desirable, therefore, that the work required to bring the testing instrument into the required position should entail the least possible distraction to the observer. Consequently, a remote control electric system has been installed to permit all movements to be controlled by the electric switchboard shown in Figs. 4 and 5. By means of push buttons the observer may cause the optical squares at A and A' to rotate to admit light to the range finder either from collimator B or from collimator B'. Another set of push buttons sets the elevation of the instrument automatically at any elevation, in steps

of five degrees, from zero to 90°, and any one of eight ranges may be chosen by a third set of push buttons.

Details of construction are shown in Figs. 6 and 7. In Fig. 6 the large 8-inch tube which forms the supporting member for the entire optical system is again shown. At the left is shown the large hook which, with the composite spring element composed of three large spring balances in parallel, provides an intermediate support to prevent sagging. The cage-like structure at the right of the center is the housing for one of the optical squares. It rotates about an axis nearly vertical for a target of zero elevation, the ribbed, *C*-shaped member carrying the upper bearing. The electric motor, below the 8-inch tube, and the crank for rotating the optical square are clearly shown. The smaller tube extending to the left above the 8-inch tube is the vacuum tube through which the light travels from one optical square to the other. It should be noted that all superstructure eccentric to the 8-inch tube is balanced by weights diametrically opposite so that no torsion is transmitted along the length of the tube. It has been the purpose of the design to have this tube at all times in equilibrium and geometrically supported.

Figure 7 shows the turret mount by which the different ranges are simulated. The disk carries eight apertures, of which one is clear and the others are fitted with achromatic wedges which introduce deviations varying from 30 seconds to 20 minutes. By these means eight different ranges can be simulated. A motor, not visible in the photograph, rotates this disk. The notches on the periphery by which the disk is indexed for the different positions are clearly shown and the elements of the group at the left of the disk are the limit switches by which the remote control is effected.

IV. INSTALLATION OF TESTING INSTRUMENT AND ACCESSORIES

The range-finder testing instrument is installed in a frame building erected specially for the purpose. The outer room contains machine tools and facilities for making modification upon range finders as required for test, desks for computational work, and suitable rooms for storage purposes and for the refrigeration machinery. The inner room, termed the test chamber, in which the testing instrument is erected, is lined with six inches of cork and provided with doors of the refrigerator type. By means of Freon refrigeration and electric heating the temperature can be maintained at any desired value between 0° and 100°F. Provision is also made for controlling the temperature to maintain a preassigned time rate of temperature rise or fall. A continuous record of the temperature in the test chamber can be made either on a circular chart or on a long roll of paper fed through a potentiometer-type recording device.

To simulate the directed radiations of the sun, a cylindrical mirror of paraboloidal cross section, in three units, extending the entire length of the range finder is provided. The heat is provided by a spiral of wire stretched lengthwise along the focal line of each unit of the cylindrical reflector. The maximum chord of the paraboloidal section is approximately 14 inches. Ordinarily, the sections of the reflector are mounted fifteen or twenty inches from the range finder and parallel to it. Under such conditions the radiant flux density, as measured at the external surface of the instrument under test, is approximately 0.8 calories per minute per square centimeter. This is approximately two-thirds of the maximum flux density that an instrument may receive when exposed to the sun at a high altitude, and experience has shown that radiant flux of this density is a powerful tool for testing a range finder, because by applying the radiant energy locally one may readily localize regions of special vulnerability.

The installation, as described, has been in use over a period of three years, and through its use much new and important information concerning range finder design has been acquired.

ACKNOWLEDGMENTS

The completion of a laboratory of this character is necessarily the result of contributions and suggestions from many people, and it seems fitting that acknowledgment should be made. A first instrument of this type adapted for one-meter beam range finders, and without provision for producing elevated targets, was designed by the writer while at Frankford Arsenal in 1922 and was successfully operated by Dr. G. W.

Moffitt. The proposal for the construction of the present laboratory was coincidental with the beginning of World War II. A first instrument of the present type was designed, under the writer's immediate supervision, by Mr. Donald H. Jacobs with the assistance of Mr. F. A. Case and Mr. Leo W. Scott. Mr. Scott supervised the design and erection of the building and the installation of refrigeration and temperature-control equipment. The early use of the testing instrument suggested so many opportunities for improvement that it has been almost entirely rebuilt. The reconstruction of the instrument was made chiefly by Mr. Robert Ward, of the Bureau machine shop, under the immediate supervision of Mr. Case with the assistance of Mr. Edgar C. Watts. Mr. Fred W. Rosberry designed several of the sub-units. The operation of the laboratory is under the supervision of Mr. L. W. Tilton, assisted by Mr. Scott.

Calendar of Meetings

August

18–20 Society of Automotive Engineers (West Coast Meeting), San Francisco, California

19–28 International Union of Geophysics, Oslo, Norway

24–27 American Institute of Electrical Engineers (Pacific General Meeting), Spokane, Washington

24–September 1 18th International Geological Congress, London, England

30–September 4 American Chemical Society (Eastern Session), Washington, D. C.

September

6–7 Mathematical Association of America, Madison, Wisconsin

6–10 American Chemical Society (Midwest Session), St. Louis, Missouri

7–10 American Mathematical Society, Madison, Wisconsin

8–9 Society of Automotive Engineers (National Tractor Meeting), Milwaukee, Wisconsin

9–11 Electron Microscope Society of America, Toronto, Canada

13–17 Instrument Society of America, Philadelphia, Pennsylvania

13–17 American Chemical Society (Western Session), Portland, Oregon

13–17 American Association for the Advancement of Science, Washington, D. C.

14–19 American Roentgen Ray Society, Chicago, Illinois

15–17 American Institute of Chemical Engineers, French Lick, Indiana

20–24 Illumination Engineering Society (National Technical Conference), Boston, Massachusetts

21–24 International Rheological Congress, Scheveningen (The Hague), Netherlands

27–29 Technical Association of the Pulp and Paper Industry, Poland Spring, Maine

27–30 American Society for Professional Geographers, Madison, Wisconsin

29–October 2 Institute of Radio Engineers, Los Angeles, California

October

5–7 American Institute of Electrical Engineers, Washington, D. C

6–9 Society of Automotive Engineers (National Aeronautic Meeting and Aircraft Engineering Display), Los Angeles, California

7–9 International Scientific Radio Union (American Section) and the Institute of Radio Engineers, Washington, D. C.

12–16 Fifth National Chemical Exposition, Chicago, Illinois

13–15 American Society of Civil Engineers (Fall Meeting), Boston, Massachusetts

13–16 Electrochemical Society, New York, New York

18–22 American Institute of Electrical Engineers (Midwest General Meeting), Milwaukee, Wisconsin

20–22 American Standards Association (Annual Meeting), New York, New York

21–23 Optical Society of America, Detroit, Michigan

24–27 American Institute of Mining and Metallurgical Engineers, El Paso, Texas

25–28 Technical Association of the Pulp and Paper Industry, Buffalo, New York

25–29 American Society for Metals, Philadelphia, Pennsylvania

25–29 Society of Motion Picture Engineers, Washington, D. C.

26 Association of Consulting Chemists and Chemical Engineers, New York, New York

27–29 Gas Discharge Conference, Brookhaven National Laboratory, Upton, Lond Island

30 American Mathematical Society, New York, New York

November

3–5 American Institute of Electrical Engineers (Southern District Meeting), Birmingham, Alabama

4–5 Society of Automotive Engineers (National Fuels and Lubricants Meeting), Tulsa, Oklahoma

4–6 Acoustical Society of America, Cleveland, Ohio

4–6 National Electronics Conference, Inc. (Annual Technical Forum), Chicago, Illinois

7–10 American Institute of Chemical Engineers, New York, New York

8–10 Institute of Radio Engineers, Rochester, New York

11–13 Geological Society of America (Annual Meeting), New York, New York

13–15 Conference on Applied Spectroscopy, Pittsburgh, Pennsylvania

The Testing of Range Finders

IRVINE C. GARDNER
National Bureau of Standards, Washington 25, D. C.
(Received January 23, 1948)

The errors of a range or height finder are divided, on the basis of origin, into four groups, namely: 1. External errors that originate from the lack of homogeneity in the air column between the instrument and target. 2. Instrumental errors that arise from defective performance of the instrument. 3. Physiological errors that are caused by physiological limits upon the performance of the observer's eye. 4. Psychological errors that have their origin in the psychological limitations of the ability of the observer to judge the range setting correctly.

Of these classes of errors, the designer can exert direct control, within limits, upon the instrumental errors. The three other classes cannot be controlled directly, but, given a qualitative and quantitative knowledge of their characteristics, it is possible to design a range finder or height finder in such manner as to lessen their detrimental effects. A brief discussion is given of the different known sources of error in each of the four classes, with typical values of the resulting errors as obtained by measurements made in the range-finder testing laboratory established at the National Bureau of Standards under the sponsorship of the Ordnance Department of the Army.

I. INTRODUCTION

THE optical range or height finder depends for its operation upon the measurement of the angular parallax of the target when viewed from the two ends of a base which is contained within the instrument. Figure 1 shows the measuring triangle of such an instrument. This triangle is nominally a right-angled triangle and is so represented in the figure. The base AB, which is subtended by the instrument is necessarily so short in comparison with the range to be measured that the angle at T, the target, is rarely as large as one-half degree. Consequently, the metrical properties of the triangle that are utilized in range finding do not vary to a significant amount whether the triangle is right-angled, isosceles, or of an intermediate shape. It may be considered that the range finder measures the angle at T, which will be denoted by α and is equal to one hundred and eighty degrees minus the sum of the angles at A and B. The fundamental equation of the range finder is

$$r = b/\alpha, \tag{1}$$

where b, referred to as the *base*, is the length AB. By differentiating

$$\Delta r = -(b/\alpha^2)\Delta\alpha = -(r^2/b)\Delta\alpha \tag{2}$$

or

$$\Delta r/r = -(r/b)\Delta\alpha. \tag{3}$$

If, in Eq. (3), one second (0.000005 radian) is accepted as a value of the probable error of the measure of the parallactic angle, it follows that the corresponding range error will be 1 percent if the range is equal to 2000 times the base length and 5 percent if equal to 10,000 times the base length. For a given base length, the percentage error in range, corresponding to a given angular error, varies directly as the range and, consequently, the linear range error varies as the square of the range. For a given range and a given angular error the range error varies inversely as the base length. Consequently, the range-finder designer is faced with two opposing requirements. Accuracy and precision requirements demand that the base length be as great as practicable but portability requirements demand a short base length. Even to approach a satisfactory compromise between these two conflicting requirements it is necessary that the design and construction be refined to permit the attainment of the maximum attainable precision and accuracy in angular measurement.

Accordingly, the optical range or height finder consists essentially of two telescopic systems designed to be used simultaneously by a single

FIG. 1. The measuring triangle of a range finder. The length AB, designated b, is the base length of the instrument. The target is at T. The range finder measures the parallactic angle α and is graduated to read range, r, directly, the graduations being based on the equation $r = b/\alpha$.

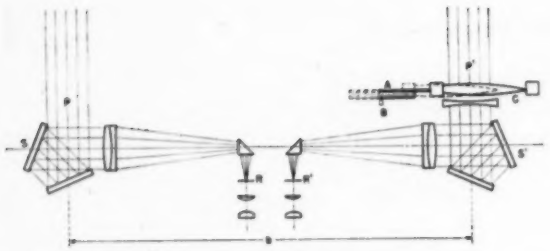


FIG. 2. Diagrammatic sketch of a stereoscopic range finder. Two beams of light from a distant target enter the instrument at P and P' and form images at the reticles R and R'. The compensator, C, consists of a negative and a positive lens which together have zero power. The positive lens may be displaced as indicated, the displacement being measured on the scale A by the index B. When the one lens is displaced the pair acts as a prism of variable deviation. The compensator is adjusted until the deviation is equal to the parallactic angle of the distant target. The scale A is graduated to indicate range directly.

observer. Figure 2 shows diagrammatically the parts of a stereoscopic range finder. The entrance pupils P and P' are widely separated to provide the required base length which is indicated by b and dimensioned on the sketch. After passing through the entrance pupils the two bundles of approximately parallel rays proceeding from the target to the two ends of the range finder are deviated toward the center of the instrument by the two optical squares S and S'. Finally, the two images of the target are imaged on the reticles R and R', respectively. The observer, looking into the binocular eyepieces, fuses the images of the two reticles into a single stereoscopic image and similarly fuses the two images of the target. In general, the target and the reticle will appear to lie in planes at different distances from the observer. The element C, placed in front of the right entrance pupil, consists of two lenses, one positive and one negative, which, when arranged axially, neutralize each other. When one lens is displaced laterally with respect to the other, as shown by the dotted lines, the converging power remains zero but a deviation is introduced. This system C, therefore, constitutes a prism of variable angle. By displacing the movable lens the apparent stereoscopic parallax is varied because of this angular deviation that is introduced, and a setting can be determined for which the image of the target and the reticle appear to lie in the same plane. Such a setting is referred to as a stereoscopic coincidence. It is evident that a scale or suitable indicating device can be attached to the movable element of system C to enable range to be read directly. Such a scale and index are indicated at A and B, respectively.

In a second form of instrument, the coincidence range finder, binocular vision is not utilized, but the two images are brought into the field of a single eyepiece. A complicated ocular prism is employed to unite the two images which appear in juxtaposition, separated by a sharp dividing line. Figure 3 shows the different types of fields as presented by typical ocular psisms. The two halves of the field proceed from the two ends of the range finder. In each pair of drawings of Fig. 3 the field to the left shows the target as it will appear, in general, through the instrument.

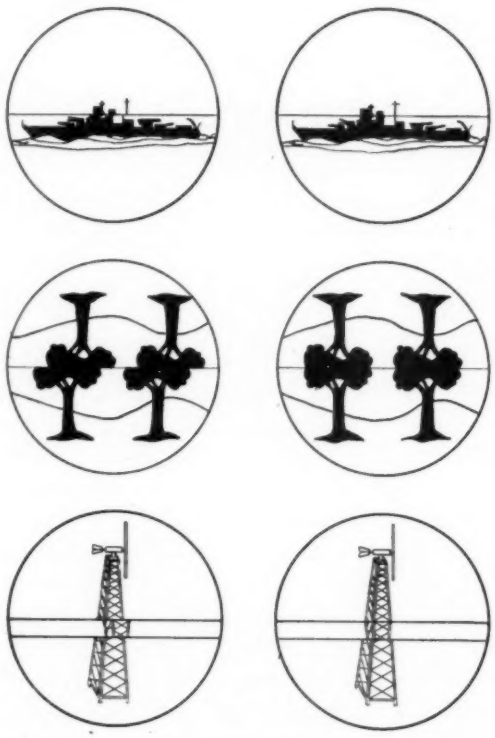


FIG. 3. Different types of fields used in coincidence-type range finders. In the upper pair of diagrams the field is divided by a horizontal diameter and the images in both halves are erect. This type is favored for instruments used against ships or other targets which present vertical lines. The central pair of diagrams shows a field in which the image in one-half is the mirror image of the other. This type is favored for use against land defenses or targets of irregular shape. With the field shown by the central pair of diagrams the upper half of the target is invisible. To remedy this defect the field shown in the lower pair of diagrams is used. A narrow horizontal strip carries the mirror image and the main part of the field above and below the strip is erect. In each case the right-hand diagram shows the appearance of the field when a correct setting is made.

The right-hand field shows the two halves of the image fitted together and is the setting for which the range is read. By using a divided field, the vernier acuity of the eye is utilized in judging coincidence, and hence greater precision is obtained than if the field were not divided and the correctness of setting judged by the simple superposition of the two images.

Range finders for infantry use have been made with base lengths as small as 20 inches. At least one range finder has been built by the British with a base length as great as 100 feet. Turret range finders on battleships have base lengths of 50 or 60 feet. Figure 4 shows a typical American height finder for use against airplanes with a base length of 13½ feet. A height finder differs from a range finder in that the slant range which is measured by the optical system is automatically multiplied, optically or mechanically, by the sine of the angle of elevation to give the vertical height of the target above the height finder. Any height finder can be used as a range finder by locking the device for introducing the angular elevation in the 90° position, for which the sine equals one. The height finder was originally devised because it was considered that the altitude of an airplane, which is the usual target, changes less rapidly than the range, and that therefore the observer could maintain the setting of a height finder on an aerial target more easily than he could the setting of a range finder. This consideration is probably still true under most circumstances, but fire control directors and other related instruments are now generally built to operate on slant range, and it is probable that the production of height finders will not be resumed. Therefore, since there are no important differences germane to the present discussion between height finders and range finders, the term range finder will be used, with the understanding that the conclusions reached are equally applicable to height finders.

Usually the ratio between the range to be measured and the base length is so large that satisfactory accuracy cannot be attained unless the parallactic angle can be measured, with an error not greater than two or three seconds.

To measure an angle in the laboratory with such an accuracy is not an easy task. It is evidently much more difficult to design an instrument meeting this requirement and also sufficiently robust and satisfactorily compensated to maintain such accuracy when transported over rough terrain in an Army truck from one battery position to another, or when subjected to the greatly varying temperatures that are encountered under service conditions.

FIG. 4. An *M*1 height finder in use by the 414th Coast Artillery Regiment at Handley Ridge, Iceland. (Photo by U. S. Army Signal Corps.)

An investigation of the extent of the available data regarding the relative magnitudes of the different errors characteristic of the range and height finders being manufactured near the beginning of World War II indicated that additional data were necessary if straightforward progress in the design of better instruments was to be achieved. As one means of meeting this requirement, a range-finder testing laboratory was established at the National Bureau of Standards, under the sponsorship of the Ordnance Department of the Army. In connection with the operation of this laboratory a new classification of the errors of a range finder was developed. This classification is now presented, with typical magnitudes for the different errors as obtained in the testing laboratory.

The errors of the range finder may be grouped as falling within four main classifications: (a) external errors, (b) internal or instrumental errors, (c) physiological errors, (d) psychological errors. The external errors are those for which the causes operate on the light before it enters the instrument. Noteworthy in this classification is the error introduced by heterogeneities in the atmosphere resulting from turbulence, with the accompanying lack of uniformity of temperature

and pressure and consequent variation in index of refraction.

The internal or instrumental errors are those that arise from imperfect instrument performance. This class of errors is subject to control as the design and construction of the instrument are improved.

Notable among the physiological errors are those that arise because of the limited capabilities of the eye considered as an optical instrument. The error resulting from the limited resolving power of the retina is an example of this type of error.

The class of psychological errors includes those that arise when the bias of the observer prevents him from correctly interpreting what is seen through the instrument.

II. EXTERNAL ERRORS

The twinkling of the stars and the irregular rapidly varying displacement of objects viewed through a column of air when conditions are favorable for convection currents are matters of common experience to all. Making a setting of a range finder, preliminary to reading range, is primarily a pointing operation, for a setting is correct when the two telescopic systems of the instrument through which the target is viewed from points A and B, the two ends of the base, are directed accurately at the target. Consequently, a measure of the error of pointing which results because of the lack of homogeneity of the atmosphere is a fundamental approach to an appraisal of the resulting error of the range finder. Dr. Washer and associates[1] have published a series of papers dealing with the different factors affecting the precision of pointing. In the course of these investigations a telescope magnifying 37 diameters was employed and the probable error of a single pointing was determined: (a) when the telescope is directed at an infinitely distant virtual image presented by a collimator in the laboratory; and (b) when the telescope is directed at a series of outdoor targets at distances ranging from 100 to 13,600 meters.

When the collimator of series (a) was employed the probable error of a single observation was found to be 0.24 second. For outdoor ranges from 100 to 13,600 meters the probable error varied from 0.46 to 0.86 second, the error increasing with the range in a fairly regular manner. This increase in the probable error is produced by the action of the long air column.

In a second investigation a range of 4710 meters, extending from the fourth-story window of a National Bureau of Standards' building to the tower of the National Soldiers' Home, was selected. The light path is 30 meters or more above the ground over substantially the entire distance, a favorable condition for a homogeneous air column. In one series of observations, pointing was effected in the usual manner, coincidence being obtained between the image of the distant target (affected by the air column) and the reticle in the telescope (not affected by air column). The probable error of a single observation was approximately 0.62 second. In a second series of observations, telephonic instructions were given by the observer to an assistant in the tower who adjusted one part of the target to bring it into coincidence with a second part. Both parts of the target to be brought into coincidence were seen through the same air path, and consequently there was no differential effect produced by the air. Under these circumstances the probable error of a single observation was only 0.26 second. Comparing these results with those of the first investigation, it is evident that the intervening air column has introduced conditions such that the probable error of a single setting has been increased from 0.26 second to 0.60 second, and this additional effect, being inherent to the air column, cannot be greatly reduced by modifications, such as increase of diameter of entrance pupil or increase of magnification, in the telescope. This concept that a portion of the probable error of setting is dependent upon the turbulence of the air is of considerable importance because it sets a minimum probable error of pointing. Experiments that have been conducted are not sufficiently extensive to fix the value of this contribution from the air definitely, but suggest that it does not exceed 0.60 second under fairly favorable conditions and for a moderate range. The importance of the effect of the atmosphere will be realized when it is noted that it is two or three

[1] F. E. Washer and H. B. Williams, J. Opt. Soc. Am. **36** 400–411 (1946). F. E. Washer and H. B. Williams, J. Research Nat. Bur. of Stand. **36**, 479–488 (1946), RP 1717. F. E. Washer, J. Research Nat. Bur. of Stand. **39**, 163 (1947), RP 1820. F. E. Washer and L. W. Scott, J. Research Nat. Bur. of Stand. **39**, 297–302 (1947), RP 1829.

times as large as the total error of pointing that has been assumed previously.

These errors in pointing would not affect the range reading of a range finder, provided that they are the same at the two ends of the base. Unfortunately, there is considerable evidence to indicate that this is not the case. The shadow bands seen at a total eclipse are evidence of the heterogeneity of the air, made apparent because the sun, immediately before the second or after the third contact, is narrowed to an approximate line source. The spacing of the shadow bands is, on occasion, eight or ten inches, indicating that the heterogeneities succeed each other after approximately this interval. Astronomers know that markings on a planet viewed with an 18- or 20-inch telescope are often better seen when the objective is stopped down to a diameter of eight or ten inches because of the heterogeneities in the atmosphere. An investigation conducted at Brown University for the Office of Scientific Research and Development and reported by Lorrin A. Riggs, Conrad G. Mueller, Clarence H. Graham, and Frederick A. Mote,[2] concerned itself with the effects of atmospheric "boil" and indicates that the displacements for two near points are essentially independent. It seems quite safe to assume that the deviations of the light entering the two ends of a meter base range finder and, more strongly, of a 13½-foot height finder, are statistically independent. From the results of Washer, it is evident that if the probable error of pointing for a single observation, for purposes of illustration, is assumed to be 0.60 second, and if the errors at the two ends are independent, the probable error of measure of the parallactic angle of a target at a reasonable range is $0.60\sqrt{2}$, which is 0.85.

This conclusion is of considerable importance and should greatly influence future range-finder design. It has often been assumed that an increase of magnification or an increase in base length of the range finder should be equally effective in increasing the precision of the instrument. Now, however, it appears that the measurement of the parallactic angle is accompanied by an accidental error, perhaps between 0.50 and 0.80 second, which originates in the air path between the observer and the target and is consequently, beyond the control of the instrument designer. By application of Eq. (3), the accidental error of range corresponding to an accidental error of 0.80 second in the measurement of α can be determined for any range and base length. Such an accidental range error arising from the atmosphere will be present even though the range finder and observer contribute no errors. Consequently, for a given range and a given precision of range measurement there is a minimum base length that is required. (See also Section IV of this article.)

During World War I it was considered a possibility that the reading of a range finder mounted on a rapidly moving ship might be systematically affected by a prismatic or lenticular mass of air compressed by the movement of the ship, which would affect the reading of a range finder by differentially deviating the light entering the two ends of the instrument. At the request of the British Admiralty, Lord Rayleigh[3] treated the subject theoretically, and concluded there would be no significant systematic error from this cause.

III. INSTRUMENTAL ERRORS

A discussion of the external errors of a range finder necessarily closes on a pessimistic note because such errors are inevitable and entirely or almost entirely beyond the control of the instrument designer. The instrumental errors, on the other hand, present a most interesting problem which is a challenge to the designer, because almost any complexity of design is justifiable, provided that it improves the accuracy and precision and provided that it can be incorporated in such a well considered construction that sturdiness and reliability of performance are not impaired. In the following discussion of instrumental errors, the errors will be classified, using as a basis the manner in which the performance of the range finder is affected and not the origin of the errors within the instrument. By adopting this procedure it is possible to avoid detailed discussion of the design and construction of the different types of range finder, a subject so broad in scope that it should desirably be the subject of separate presentation.

[2] L. A. Riggs, C. G. Mueller, C. H. Graham, and F. A. Mote, J. Opt. Soc. Am. **37**, 415–420 (1947).

[3] Lord Rayleigh, Trans. Opt. Soc. (London) **20**, 125–134 (1918–1919).

A. Errors that are primarily a function of range (calibration errors).—Most measuring instruments of any pretense to accuracy are provided with an adjustment by which the scale is caused to indicate the correct value at one or more graduations. A homely example is the ordinary platform scale, which is customarily adjusted to read zero when there is no load. The corresponding adjustment of a range finder is termed an "infinity" adjustment, because the value for which the adjustment is made corresponds to zero parallactic angle, that is, to infinite range. By a combination of optical and mechanical devices the scale of a range finder is made very open, one-tenth of an inch corresponding to approximately one second. It is self-evident that the scale can be so graduated that at some one temperature all graduations will be correct after the infinity adjustment has been properly made. In a laboratory instrument, if this were not the case a calibration correction could be applied. The conditions of use of a range finder do not permit calibration correction, so production is necessarily controlled until the required accuracy throughout the length of the scale is assured. However, there is a more subtle error of this type under discussion. Even if the scale is correctly graduated for one temperature it cannot be assumed, without consideration, that the infinity adjustment at some other temperature will suffice to insure correct readings over the whole scale. The importance of this question becomes more significant when it is recalled that an error of reading corresponding to one second is readily appreciated on any part of the scale of a range finder, and also that a range finder is supposed to function satisfactorily over a very large temperature range, say −50°F to 125°F. Actually both theory and experience indicate that the calibration errors can be kept small at all temperatures. In substantially all range finders the optical and mechanical parts are so constructed that a scale, correct at one temperature, performs satisfactorily at any temperature when adjusted to read correctly at one graduation, and, with reasonable care, this satisfactory behavior may be expected to persist throughout the life of the instrument. Before dismissing the subject it should be mentioned that erroneous conclusions may be reached if one attempts to check the correctness of the scale throughout its length by checking it against a variety of terrestrial targets at different ranges, because there are psychological errors which affect the different readings differently. This will be discussed in the section dealing with psychological errors.

B. Errors that are a function of the angular elevation of the line of sight (elevation errors).—It is a matter of experience that a range finder adjusted to read correctly when directed horizontally at an object may be subject to very large errors when directed at an object overhead. For a naval range finder used against opposing ships this is not a serious error, but for a range finder used against airplanes such an error is a very important one, because, even for a single target, during the period of approach the angle of elevation varies over a very large range.

To direct a height finder at an elevated target, the entire instrument is rotated about its longitudinal axis and the orientation of the entire optical system is changed with respect to the gravitational field. Consequently, one may expect the alignment of the optical parts and the shapes of the mechanical parts to be different for different elevations. In an instrument reading to seconds, changes of this nature might be expected to be significant. For the earlier types of range finder this undoubtedly occurred. More modern types of instruments, however, are designed to make displacements of this character self-compensating. Consequently, errors arising because of alteration of mechanical strains as the line of sight is elevated are relatively unimportant in a modern range finder.

An elevation error may arise because of temperature gradients in the air in the interior of the range finder. Referring to Fig. 1, the path AB, the base length, is enclosed within the instrument. It is possible, and experience shows it is also probable, particularly when the instrument is exposed to directed radiation, that a vertical temperature gradient may exist in this air path. The path of the light in the instrument is normal to the gradient, and therefore the ray path, instead of being a straight line, is an approximately circular arc. If 1 (measured in centimeters) is the path length, dt/dx is the temperature gradient (degrees centigrade per centimeter), and dn/dt (change in index of refraction per degree centi-

grade), the total angular deviation, D (measured in seconds), of the ray is given by the equation

$$D = 2.06l(dt/dx)(dn/dt)\cdot 10^5.$$

For air, dn/dt is approximately 1×10^{-6}. If l is 400 centimeters and dt/dx is assumed to be 0.1, the deviation is 8 seconds. This deviation is in a vertical plane. Consequently, when the angle of elevation is small, the measuring triangle TAB is nearly horizontal, the deviation is approximately normal to the plane of the triangle and no significant error is introduced. This source of error, therefore, remained unimportant until anti-aircraft fire assumed a prominent role in warfare. When the instrument is directed at an elevated target, such as an airplane, a greater component of this error lies in the plane of the measuring triangle and introduces a significant error. In particular, for the assumed values the error is 8 seconds for 90° elevation. This is a very important source of error, because dt/dx, under unfavorable conditions, may have values two or three times that assumed in the example.

There are two methods of ameliorating this error. If the instrument is filled with helium instead of air, the value of dn/dt is approximately one-eighth that for air, with consequent reduction in the error. There is a further reduction, because helium, by reason of its greater conductivity, also causes a reduction in dt/dx. It is also possible to reduce the value of dn/dt to any desired value by evacuating the interior of the instrument.

As a historical digression, it is interesting to note that the firm of Zeiss was aware of this source of error as early as 1934 and avoided it by evacuation. In 1936, helium, obtained as a by-product by the firms extracting argon and neon from the air, was substituted for the vacuum in order to avoid the pressure difference. The British discovered this error during the course of World War II, referring to it as "stratification," and communicated it to this country. The application of helium in this country was an independent discovery, and it was not known that the same means had been employed in Germany until captured range finders were examined at the National Bureau of Standards. The delay in discovering this very large source of error, of importance only when the range finder or height

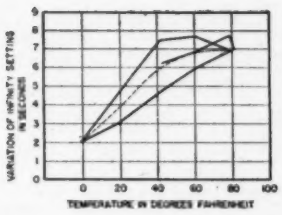


FIG. 5. Hysteresis of a height finder as the instrument is subjected to a closed cycle of equilibrium temperatures. The range of temperature extends from 0 to 80°F. From the diagram it is evident that readings at 40°F may be expected to differ by as much as three seconds when in temperature equilibrium, the difference being dependent upon whether the equilibrium is arrived at from a higher or lower temperature. Experience shows that these hysteresis curves are characteristic, even though successive runs do not always repeat exactly, as is shown by the failure of this curve to close. The dotted line is a median curve of the cycle and shows the shift in the infinity adjustment with temperature if the hysteresis effect is excluded.

finder is directed at an elevated target, is presumably due to the practice of testing such instruments for acceptance only on targets at low elevations, because natural fixed targets of known range, five to ten thousand yards distant and at elevations of 45° or more, are not readily available.

C. Errors that are introduced when a range finder is changed from one equilibrium temperature to a second equilibrium temperature.—In the discussion of errors which are a function of angular elevation, mention has been made of errors resulting from mechanical strains in the parts of the instrument. It is obvious that strains due to thermal changes may also prove to be a source of error. In considering thermal effects, the simplest case is presented when the instrument is studied at a series of temperatures when in thermal equilibrium. It has been found interesting to choose the equilibrium temperatures to constitute a cycle, beginning, for example, at 60°F, proceeding by steps of approximately 20° to 100°F, then decreasing to 0°F, and returning to 60° or perhaps 80°F, and to make measurements of the consequent shift of the infinity setting for targets at 0° and 90° elevation. After each change of temperature it is necessary to wait 14 to 16 hours to make certain that a satisfactory approximation to thermal equilibrium has been attained. The graph of Fig. 5 illustrates the performance of an instrument which is not of best quality, as regards this particular criterion. The abscissae denote temperatures, and the ordinates are the

corresponding shifts of the infinity setting measured in seconds of arc from an arbitrary zero. The cycle of temperature yields an open curve suggestive of the hysteresis curve of a magnetic material and indicates that the readings of this particular instrument, when in thermal equilibrium at a given temperature, may differ by several seconds, the value depending upon whether the equilibrium temperature has been attained from a higher or from a lower temperature.

If all parts of a range finder are homogeneous and well annealed, at two equilibrium temperatures the shapes of a given part will be geometrically similar, provided that mechanical strains are relieved, and a study of the details of design suggests that one should expect little or no hysteresis effect. However, frictional forces may exist between adjacent parts, and this may cause the retention of mechanical strain when the instrument passes from one equilibrium temperature to a second. Such an assumption is sufficient to account for the hysteresis of Fig. 5, as the retained mechanical strains will differ in sense as equilibrium is approached from the two directions.

In Fig. 5, a dotted line is drawn through the average values of the two ordinates for each temperature. The slope of this curve is the temperature coefficient of the infinity adjustment when temperature equilibrium is maintained. The performances of individual instruments vary greatly, some showing a temperature coefficient nearly zero, with large amounts of hysteresis, and others showing a larger temperature coefficient and less hysteresis. In the newer instruments, when optical compensation to eliminate the effect of mechanical strain is relatively perfect, the hysteresis is small and this accords with the suggestion that it is due to mechanical strain that persists because of frictional restraints.

D. Errors that are introduced when the ambient temperature changes so rapidly that the instrument does not remain in temperature equilibrium.—It has been mentioned that if all parts of a range finder are well annealed, their shapes, at different equilibrium temperatures, will be geometrically similar, angular configurations will be unchanged, and the consequent error might be expected to be small. When the ambient temperature changes rapidly, however, there will necessarily be temperature gradients in the instrument, with consequent deformation of the parts, and large contributions of error are to be expected. Errors of this character are large, and it is difficult to control conditions sufficiently to have them repeat closely, because, in general, they are a result of the history of the instrument as well as of its present state. This will be the more readily understood when the hypothesis introduced to explain hysteresis is recalled. It was assumed that mechanical parts, when changing their dimensions as a consequence of passing from one equilibrium temperature to another, were prevented from changing freely by frictional forces which caused mechanical strains to be built up of which a portion is retained. During the period of changing ambient temperatures it may be expected that these strains will build up to certain values, at which time the frictional forces will be overcome. There will be a sudden lessening of strain, and then the process will be repeated, giving rise to a "chattering" type of change. If the type of compensation desired to give a range finder thermal stability is entirely optical, both hysteresis and the errors arising during changing temperatures will be less than when the thermal stability is affected by the behaviors of mechanical parts. Errors arising from the deformation of mechanical parts can be decreased by supporting all vital mechanical parts geometrically to permit them to respond fully and promptly to thermal changes. Even when this is done, however, trouble can still arise because of the presence of thermal gradients. Errors from changing ambient temperatures are often large, amounting to three or more seconds for changes at the rate of 20°F per hour. Optical stabilization, careful design and geometric support of important mechanical parts, and the reduction of temperature gradients by the use of material having high thermal conductivity lessen the errors caused by changing temperature. The diurnal change of temperature is a universal stimulus for errors due to changing temperature. To reduce or eliminate this source of error, it has been proposed to thermostat the instrument to maintain it continuously at a constant temperature. Such a measure should be effective, but it is evidently a difficult one to apply, more particularly for instruments used by the army in temporary installations.

E. Errors that occur when a range finder is exposed to directed radiant energy.—Rapidly changing ambient temperature and directed radiant energy probably produce errors in range-finder performance through the same ultimate agency, namely, the production of temperature gradients within the instrument. In the tests made at the National Bureau of Standards the directed radiation was produced by a series of electrical heaters distributed along the focal line of a cylindrical reflector of parabolic cross section. The length of the reflectors is approximately equal to that of the range finder. The dimensions and heating elements are such that the radiant flux is approximately equal to two-thirds that which is received by an instrument exposed to the summer sun. The effects produced by radiant flux of this magnitude are greater than those produced by the standard rate of change of ambient temperature that has been adopted for test purposes. Errors arising from radiation often amount to 6 to 10 seconds.

When the directed radiant flux is not applied along the entire length of the instrument but only to selected parts, the response for different portions of the length differs greatly, and such procedure enables the source of the error to be localized. For the older type of range finder, not stabilized with respect to temperature, the central part of the instrument is sensitive to directed radiant flux and it is not unlikely that the error is produced by the bending of the optical bar resulting from transverse temperature gradients. Such an error could occur even if the range finder were evacuated or filled with helium. The range finders stabilized against equilibrium temperature changes and evacuated or filled with helium are often sensitive to exposure to radiant flux at the ends, because usually there are sections at either end that are filled with air and hence poorly protected against temperature gradients and consequent stratification. Optical squares are also an important cause of sensitivity at the ends, because temperature gradients often cause a change in the angle between the two reflecting surfaces and hence change the deviation appreciably.

A length of two or three feet at each end of the German R40 range finder was protected by a radiation shield of thin metal spaced approximately one inch from the body of the instrument.

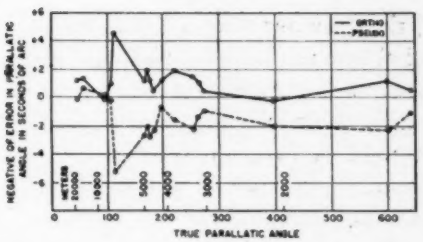


FIG. 6. Graph showing psychological bias. The two curves represent readings with the same range finder when adjusted to give ortho- and pseudostereoscopic vision. The extent to which one curve is the mirror image of the other illustrates the error, measured in seconds of arc, resulting from psychological bias. All the targets were terrestrial natural objects, and the irregularity of the curves shows that the amount of the bias varies greatly for different targets.

Tests made with and without these shields in position proved their effectiveness. The R40 instrument was one of the most modern German range finders in which the greater part of the significant portion of the interior was filled with helium. The body of this instrument was of thick cast aluminum which, by reason of its high thermal conductivity, probably lessened the effect of directed radiation over the entire interior of the instrument.

IV. PHYSIOLOGICAL ERRORS

Originally the limited resolving power and limited acuity of the eye were considered to introduce important limitations upon the performance of a range finder. To lessen errors of this nature, high magnifications were employed and, with a given type of range finder, it was tacitly assumed that the accidental error in measurement of the parallactic angle in the true field of a range finder would vary inversely as the magnification. When tests demonstrated that this was not the case, sources of error were sought elsewhere and, as a result of Dr. Washer's work, it is established that a probable error somewhat less than one second is to be expected because of the turbulence of the atmosphere. Under favorable conditions, with a magnification of 12, an observer with normal vision can attain this precision. This does not necessarily imply that higher powers should not be used, because, with the higher powers, the observer may be enabled to make successive settings rapidly with confidence or with less fatigue. On the other hand, for types of work in which a wide field is very de-

sirable, the designer may use a lower magnification and obtain the wide field without feeling that the demand for precision has been entirely abandoned. Physiological errors play an important part in connection with the design of range finders for use at night and for other unusual uses, but the study of the external errors of range finding has definitely lessened their importance as a determining factor in range-finder performance.

V. PSYCHOLOGICAL ERRORS

It requires only a limited observation of stereoscopic vision to learn that the stereoscopic judgment is one that is easily biased and yields incorrect judgments because of characteristics, other than stereoscopic parallax, of the object under observation. In ordinary life the perception of depth based on stereoscopic parallax is generally aided, and at times entirely supplanted, by other contributing evidence, such as perspective, shadows, and the so-called atmospheric perspective or blue haze effects which increase with distance. In making a setting with a stereoscopic range finder the reticle and target appear to come into coincidence as the range knob is turned. If the target is silhouetted against the sky, many observers find it desirable to set the reticle mark at one side of the target, and, for a careful setting, coincidence between target and reticle is disestablished alternately in opposite senses until the proper setting is reached. If, however, the target is extended or against a background so that there is no free space on either side of the target, then when the reticle is projected beyond the surface of the solid object the observer is loath to accept the apparent penetration of the reticle into the solid target, and hence a bias is introduced which may correspond to a large error in the indicated range. Similarly, when one attempts to set a crisp, sharply defined image of the reticle in stereoscopic coincidence with a target not sharply defined and obscured by blue haze, the pronounced difference in atmospheric perspective leads to an erroneous setting. These errors resulting from biased stereoscopic judgment may amount to as much as five or six seconds. If one attempts to check the graduations on the scale of a range finder by reading ranges on a selected series of terrestrial targets at known distances, the stereoscopic bias for the different targets will, in general, be different for the different targets, giving a jagged calibration curve which suggests that individual graduations on the scale are displaced from their correct positions in a non-uniform manner. The falsity of this conclusion can be shown by testing the range finder with a better planned test in which stereoscopic bias is eliminated or equalized for the different targets.

The type R40 German range finder has a set of prisms which can be introduced into the optical system to change from ortho to pseudostereoscopic vision. In pseudostereoscopic vision the rays from the target entering the left entrance pupil are transmitted to the right ocular and *vice versa*. The effects of stereoscopic bias on the measured range for ortho and pseudostereoscopic vision are equal in magnitude but opposite in sign. Consequently, if calibration curves on terrestrial targets are made with ortho and pseudostereoscopic vision, and the two curves plotted on a common system of coordinates with error as ordinate and range as abscissa, the one curve should be obtained by mirroring the other in the axis of abscissae. Figure 6 is an example of such a pair of curves. The striking manner in which one curve is the mirror image of the other gives impressive confirmation to the importance of psychological bias as affecting the accuracy of the stereoscopic range finder. The targets used were terrestrial objects at the different distances, and the irregular course of each curve is clearly due to psychological bias which is different for the different targets, in one case amounting to somewhat more than four seconds. If a range finder can be used with either ortho- or pseudostereoscopic vision the average of ortho- and pseudo-readings may be expected to eliminate this error, but this method can be applied only to stationary or slowly moving targets. Consequently, the provision of prisms to permit either ortho- or pseudostereoscopic vision to be realized in a range finder is a valuable aid for training purposes or for range reading on stationary or slow moving targets. Experiments on the effect of psychological bias on coincidence-type range finders have not been made, but there are *a priori* reasons for believing that this error will be less than for the stereoscopic instrument.

A Method for the Evaporation of Alloys*†

LOUIS HARRIS
Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

BENJAMIN M. SIEGEL**
Weizmann Institute of Science, Rehovoth, Palestine

(Received January 30, 1948)

We have developed a method for the evaporation of alloys in high vacuum which from the nature of the procedure used should be applicable to all alloys which are not refractory. The method has been tested on α-brass and β-brass and on a gold-cadmium alloy. These alloys have components whose vapor pressures are very different and yet chemical and electron diffraction analyses on the evaporated brass alloys and chemical analysis on the evaporated cadmium-gold alloy show that the composition of the original alloy is held to within close limits.

INTRODUCTION

THE evaporation of metal alloys in a high vacuum is attended by several difficulties. For example, when an alloy is heated to the temperature required for evaporation, the vapor phase has a composition different from the original solid phase and the composition of the vapor phase will, in general, change during the course of an evaporation. An added difficulty may be the selective "reflectance" of the components at the condensing surface for some alloys. Thus, the condensed alloy obtained from the evaporation of an alloy sample may be quite different in composition from the original one. Undoubtedly, it would be possible to determine the composition of an alloy, which upon evaporation would produce an alloy of the desired composition in the condensed state. Another possibility would be to determine the conditions under which two pure metals would have to be evaporated in order to produce a vapor which would condense to give the desired binary alloy. Both of these methods, however, require such a large amount of empirical data as to make the methods impracticable, except in special cases. We have developed a method for the evaporation of alloys which has proved successful and which promises to be generally applicable to all alloys whose components can be evaporated.

The procedure used in our method is to evaporate a few finely divided particles of the given alloy almost instantaneously and continue this process until sufficient alloy has been deposited. We accomplish this by dropping the powdered alloy, a few grains at a time, into a refractory boat heated well above the temperature required to evaporate *completely* the components comprising the alloy. Thus, as each particle falls into the hot boat, it is evaporated completely, and, as successive particles continually drop into the boat, a vapor phase is established which has the same composition as that of the original alloy.

For our experiments we have deliberately chosen binary alloys in which the vapor pressures, at all temperatures, for the two components are widely different. The alloys chosen were from the copper-zinc system and a gold-cadmium alloy.

EXPERIMENTAL PROCEDURE

The apparatus used for the study of the evaporation of alloys is shown in Fig. 1. The complete array is contained in a bell jar which is evacuated below 10^{-4} mm of mercury for the evaporation. The continuous belt on which the powdered alloy is placed and the spring mechanism for moving the belt are shown. The belt motion is started by releasing a holding pin by the activation of a solenoid, from outside the

* This work has been supported by the Navy Department under Contract NObsr 25391; the preliminary part of this work was done under National Defense Committee, research contract NDCrc 180.

† Contribution from the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts.

** At the Polytechnic Institute of Brooklyn by special arrangement.

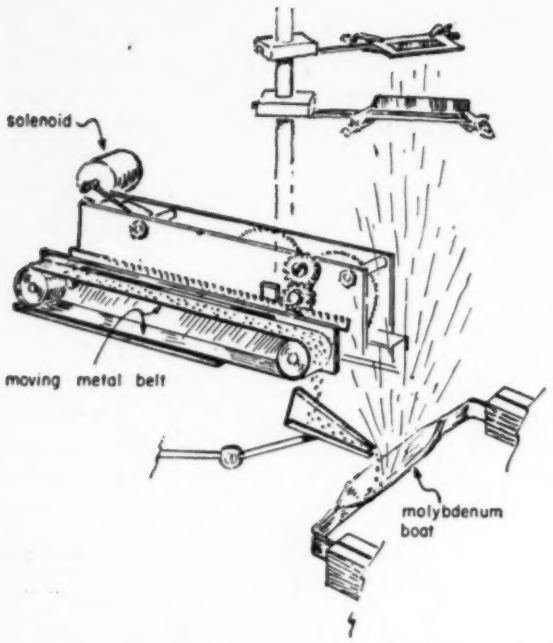


FIG. 1. Apparatus for evaporation of alloys.

bell jar. As the belt moves, the sparsely dispersed powdered alloy drops, a few grains at a time, into the boat, which in these experiments, was made of molybdenum sheet. Powders of 80 to 100 mesh gave the optimum size for alloys such as brass. The temperature of the boat was adjusted so that the particles upon falling into the boat could be seen to melt and then disappear almost instantaneously. The temperature of the boat must not be too high, else many of the particles may be "blown" from the boat without having undergone complete evaporation. Under such conditions, the condensed alloy was found to have a composition different from that of the original alloy.

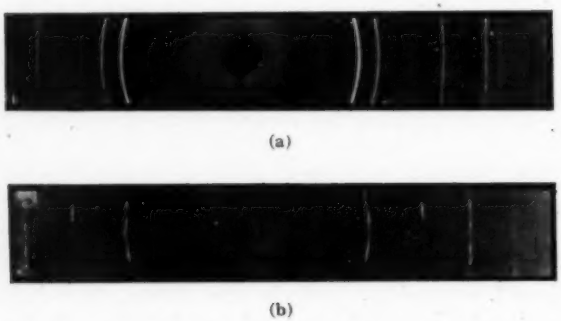


FIG. 2. X-ray diffraction patterns of powdered samples of α- and β-brass used for evaporation.

METHOD OF ANALYSIS

The composition of the original and of the condensed deposits was obtained by chemical analysis. Since the deposits weighed only a few milligrams it was necessary to use a micro method of analysis. We used a micro-colorimeter method involving the reagent "dithizone"[1] and used a "photelometer" to evaluate the transmissions of the solutions used. After some modifications an accuracy of ~2 percent was obtained.

The crystal structures of some of the condensed deposits were investigated by electron diffraction studies. Crystallite size on some of the deposits was obtained with the electron microscope.

RESULTS

Gold-Cadmium Alloy

A specially prepared gold-cadmium alloy, 63.8 percent by weight of gold and 36.2 percent by weight of cadmium (50 atomic percent gold-cadmium) was evaporated by the above procedure. The deposits obtained upon evaporation had the same chemical composition as the original alloy.

No attempt was made to check the crystal structure of the evaporated deposit by electron or x-ray diffraction.

α- and β-Brass

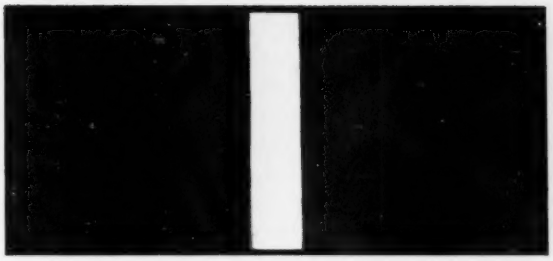
Chemical analysis of a deposit obtained from the evaporation of powdered β-brass upon a glass slide showed the same composition, 51.7 percent copper and 48.3 percent zinc, by weight, as the original β-brass sample.

Specimens of α- and β-brass (no chemical analyses made for these samples) were deposited by evaporation upon collodion films (film thickness ~500A) so that transmission electron diffraction and electron microscope studies could be made. The powdered alloys were examined by x-ray diffraction to ascertain their crystal structure and composition and the patterns obtained are given in Fig. 2. The α-brass sample shows the augmented spacings of the face-centered cubic copper lattice, characteristic of this alloy. The β-brass sample gives only the body-centered cubic lattice lines of copper:zinc, 1:1. Figures 3a

[1] E. B. Sandell, *Colorimetric Determination of Traces of Metals* (Interscience Publishers, New York, New York, 1944).

(a) (b)

FIG. 3. Electron diffraction patterns of samples obtained from the evaporation of α- and of β-brass.



(a) (b)

FIG. 4. Electron microscope photographs of samples obtained from the evaporation of α- and of β-brass.

and 3b give the electron diffraction patterns obtained from the two *evaporated* samples of α- and β-brass, respectively. Figure 3a is again the diffraction pattern obtained from a face-centered cubic lattice and the crystal spacings are the same as those given by the x-ray data, on the original alloy. Figure 3b shows very clearly the body-centered cubic lattice and the lines in the pattern are those corresponding to β-brass alone. Since the β-brass structure is to be found as a single phase, at room temperature, only between 47 and 50 weight percent of zinc, the alloy deposited by evaporation must be correct to within that range.

The electron microscope results are shown in Fig. 4 and it can be seen that deposits appear discontinuous at a magnification of 50,000×. The initial crystallite formation is of the order of 100–200A for β-brass, but in α-brass the crystallites grow slowly to about that size. The broadening of the diffraction rings observed for α-brass taken shortly after evaporation correspond to crystallites of only 30–50A.

DISCUSSION OF RESULTS

Both cadmium and zinc are notoriously difficult to condense on cellulose nitrate films at room temperature. It has been demonstrated[2] that zinc (cadmium) will condense readily on copper (gold) deposits. Such a condition is realized very soon after evaporation has begun, in our experiments. The chemical analyses show that the "reflection" of the vapor stream from the condensed deposit is non-selective in the components, for the evaporation technique used here.

Since we did not obtain electron diffraction results of the brass deposits immediately upon completion of evaporation, we cannot be certain of the initial lattice arrangement of those deposits. König[3] has demonstrated that diffusion is comparatively rapid when zinc is evaporated over evaporated copper deposits or when copper is evaporated over evaporated zinc deposits. However, a homogeneous deposit was obtained in very thin layers deposited here, as shown by the electron micrographs and corresponding electron diffraction patterns. Thus, though the vapor pressure of each of the components in the alloys was very different, the alloys must have condensed from a vapor phase of the correct composition to give small crystallites in which each have the proper proportions. Any alloy in which the components may be expected to diffuse even slightly into one another can therefore be expected to produce a homogeneous deposit by this method of evaporation.

We believe that we have described a procedure useful for the producing of alloys in thin layers. The method should prove useful in the production of special photo-emissive surfaces, of thin strips of alloys having special electrical properties such as abnormal temperature coefficients of resistivity, and in fundamental studies of the properties of alloys.

ACKNOWLEDGMENTS

We are indebted to Mr. Willard E. Buck for his assistance and suggestions in initiating this work, to Miss Margaret C. Hubbard and Mrs. Betty K. Lyon for their assistance with the chemical analyses, and to Mrs. Rose Brenner for the x-ray diffraction examination of the powdered samples.

[2] E. Mollivo, Reichsber. für Physik 1, 1 (1944).

[3] H. König, Reichsber. für Physik 1, 7 (1944).

Statistical Factors Affecting the Intensity of X-Rays Diffracted by Crystalline Powders*

LEROY ALEXANDER, HAROLD P. KLUG, AND ELIZABETH KUMMER
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh 13, Pennsylvania

(Received February 2, 1948)

The Laplacian probability equation can be used to predict the reproducibility of the intensities of x-rays diffracted by crystalline powders. If absorption is neglected, and N is the number of particles irradiated, p the proportion of particles which are oriented so as to diffract, and $q=1-p$, the relative mean deviation to be expected in repeated intensity measurements is given by $U_m=0.798(q/Np)^{\frac{1}{2}}$. When absorption and the actual distribution of particle sizes are taken into account, the expression becomes

$$U_m=0.798(q\mu v_e/pA)^{\frac{1}{2}},$$

where μ is the linear absorption coefficient of the material, v_e the effective volume per particle, and A the cross-sectional area of the incident beam.

The theory is confirmed by measurements on four size fractions of quartz powder using a Norelco Geiger-counter x-ray spectrometer. For typical experimental conditions and powders of medium absorbing power for x-rays, it is demonstrated that good reproducibility can be expected only if the effective particle dimension is less than approximately five microns.

I. INTRODUCTION

SINCE the advent of the science of x-ray diffraction the importance of intensity measurements from crystalline powders has been recognized. Such measurements have been valuable in the experimental determination of atomic scattering factors. These fundamental quantities, of course, have formed the basis for all subsequent structure analysis, where, likewise, precise intensity measurements have been indispensable both when single crystals and powders have been employed. Indeed, the most accurate intensity data are often none too good to permit the selection of the true atomic arrangement from among the erroneous alternatives. The intensities of powder "reflections" have also been of invaluable assistance in the study of extinction phenomena. More recently there have arisen extensive applications of powder diffraction methods in the field of quantitative analysis of mixtures. Here again the value of the method is limited largely by the accuracy of the intensity measurements.

For the ideal case of a very large number of particles randomly oriented, the ratio of the power of the recorded portion of the diffracted beam to that of the primary beam is[1]

$$P_d/P_0=QmL\delta V/8\pi rA\sin\theta. \qquad (1)$$

In this expression Q has its usual significance and depends upon the structure factor of the planes contributing to the reflection, the wave-length of the x-rays, the number of unit cells per unit volume of crystal, and the Bragg angle. δV is the volume of powder irradiated and is here assumed to be sufficiently small that absorption may be neglected. m is the multiplicity factor for the set of planes concerned, while L, r, and A depend upon the geometrical features of the experimental arrangement.

For a reflection from a given set of planes (hkl) and under a fixed set of geometrical conditions, it is seen from Eq. (1) that the ratio P_d/P_0 is proportional to δV, the effective volume of powder (absorption being negligible). Equation (1), of course, presupposes the absence of primary and secondary extinction effects, a condition which may be closely approached in a good powder specimen.

For any actual powder specimen the total number of particles[2] is always finite and may

* Based on material presented before the American Society for X-Ray and Electron Diffraction at Ste. Marguerite Station, Québec, June 25, 1947.

[1] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Company, Inc., New York, New York, 1935), second ed., p. 417.

[2] Throughout this paper the term *particle* is construed to mean an individual *crystallite*, never a polycrystalline aggregate.

often not be sufficiently large to guarantee the complete applicability of Eq. (1), even assuming that extinction effects are negligibly small. If N is the total number of particles irradiated by the incident beam, and p is the ideal fraction which is oriented so as to give rise to a diffracted ray from the planes (hkl), the number of particles theoretically contributing to the intensity of the diffracted ray is Np. If absorption can be neglected (or is uniform for all the particles) and the particles are of equal size, the intensity of the diffracted ray is proportional to Np.

The ideal proportion of favorable orientations, p, depends upon the crystallite perfection and the particular geometrical conditions under which diffraction takes place and the diffracted rays are recorded. Therefore, for any actual experimental arrangement p has a definite and fixed value. Probability theory predicts, however, that the number of particles actually oriented so as to contribute to the reflection (hkl) will not be precisely Np but will differ from it by a small amount u. Assuming that the particles are random in shape and that the orientation of any given particle is independent of the orientations of the others, the probability of any given deviation u is predicted by the theory.

If we represent the actual number of favorably oriented particles by n_f, u is defined by:

$$u = n_f - Np. \tag{2}$$

Laplace's generalized probability formula[3] states that the probability of a deviation of magnitude u is

$$P_u = (2\pi Npq)^{-\frac{1}{2}} \exp(-u^2/2Npq), \tag{3}$$

in which $q = 1 - p$ and represents the theoretical fraction of unfavorable orientations. This formula is sufficiently accurate only if u is very small in relation to both Np and Nq. If u is only moderately smaller than Np and Nq, a more accurate value of the probability can be obtained from the formula

$$P_u = (2\pi Npq)^{-\frac{1}{2}} \exp[-(u^2+u)/2Npq]. \tag{4}$$

When Eq. (3) is applicable, the standard, probable, and mean deviations are given respectively by the following expressions:

$$\begin{aligned} \sigma_s &= (Npq)^{\frac{1}{2}}, \\ \sigma_p &= 0.675(Npq)^{\frac{1}{2}}, \\ \sigma_m &= 0.798(Npq)^{\frac{1}{2}}. \end{aligned} \tag{5}$$

The relative deviations are obtained from the absolute deviations by dividing them by the number of favorable orientations, Np. So we obtain:

$$\begin{aligned} U_s &= (q/Np)^{\frac{1}{2}}, \\ U_p &= 0.675(q/Np)^{\frac{1}{2}}, \\ U_m &= 0.798(q/Np)^{\frac{1}{2}}. \end{aligned} \tag{6}$$

These expressions make it possible to predict the probable or mean deviation in the actual number of favorably oriented crystallites from the ideal number, provided N and p are known. Since the intensity of the diffracted ray is dependent upon the quantity Np, it follows that we are also in a position to predict the degree of reproducibility to be expected in the intensities of x-rays diffracted from powders of known particle size. From Eqs. (6) it is seen that if N and p are sufficiently large, the expected relative deviations are small; that is, intensity measurements from successive samples of the same lot of powder should show good agreement. As N and p diminish in size, the reproducibility will deteriorate until a point is reached beyond which a single measurement will not yield a reliable result.

In the present communication the extent of these effects will be investigated for the particular case of diffraction from the surface of a stationary flat powder cake, utilizing an asymmetric focusing arrangement. This set of experimental conditions is favored at the present time for obtaining highest quality intensity data from powders. A particular advantage of the arrangement is the ready evaluation of absorption effects which it permits.

It will be assumed in the following discussion that the focal spot of the x-ray tube itself acts as the source of diverging x-rays, the beam being observed at a small angle to the face of the target so that a vertical line focus is approximated. A single slit is placed between focus and sample to limit the beam. It will be further assumed that the diffracted rays are received by a narrow slit, the receiving slit being rotated about the sample,

[3] See R. B. Lindsay, *Introduction to Physical Statistics* (John Wiley and Sons, Inc., New York, New York, 1941), pp. 25–26, or any other standard work on the subject.

and the sample being rotated so that its surface makes an angle θ with the incident beam when the receiving slit is located so as to receive a beam at the angle 2θ. A further requirement for optimum focusing is that the x-ray source-to-sample distance be equal to the sample-to-receiving slit distance. The source, sample, and receiving slit, then, all lie at points on a focusing circle of radius $L/2\sin\theta$, where L is the sample-to-source and sample-to-receiver distance.

II. PROPORTION OF PARTICLES FAVORABLY ORIENTED

As emphasized in Section I only two quantities, N and p, must be evaluated for any given experiment in order to predict the degree of reproducibility of intensity measurements. Of these quantities p is inherently the more difficult to determine. Although a rigid analysis of all the experimental conditions does not permit a precise evaluation of p, its value can be shown to lie in such a definite range that valid predictions are possible.

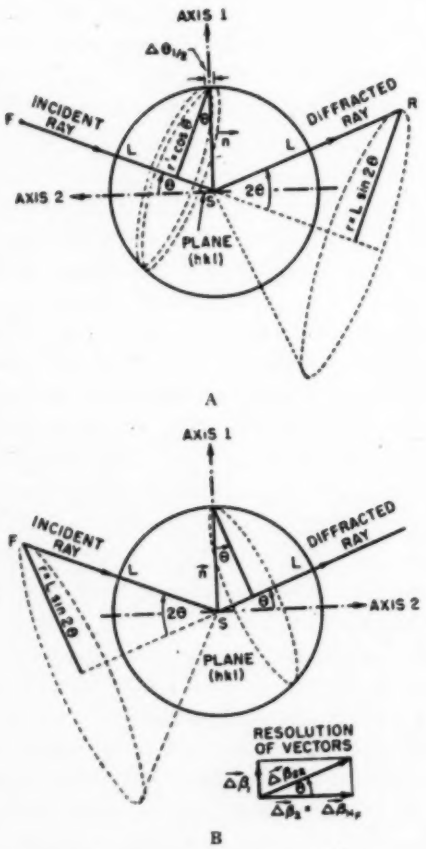


FIG. 1. Relationships of diffraction to the geometry of the apparatus, and representation of crystal orientations by means of an orientation sphere.

The size of p is a function of (1) the rocking angle of the crystallites constituting the powder and (2) the geometrical conditions under which diffraction takes place and the diffracted energy is received. The rocking angle of a crystal is a measure of its perfection. It can be defined as the angular range, $\Delta\theta_i$, through which the crystal can be rotated without reducing the intensity of the diffracted beam to less than one-half the maximum intensity (intensity diffracted at the Bragg angle θ). For ideally perfect crystals $\Delta\theta_i$ is only a few seconds of arc while for very imperfect crystals it may amount to as much as one degree of arc. Thus, specimens of calcite have been found which possess a rocking angle for the first-order cleavage face reflection of only 4.4″ using AgKα radiation.[4] Haworth and Bozorth[5] have found that single quartz crystals may show an even higher perfection, reflections from etched $(11\bar{2}0)$ faces displaying rocking angles as low as 2.6″ with MoKα radiation. Most crystals are relatively imperfect and possess rocking angles measured in minutes of arc.[5] For example, reported rocking angles for the (100) reflection of rock salt vary from 174″ to 30′ depending upon the degree of imperfection of the crystal examined.[6,7,8] For fine powders prepared by crushing and grinding of massive crystals, $\Delta\theta_i$ is necessarily much larger than the theoretical value for ideal crystals because of the imperfections introduced by the mechanical treatment. Quartz powder crystallites might be expected to possess a mean rocking angle about equal to that of average single crystals of rock salt (5′ to 20′).

From the standpoint of the diffraction effects the orientation of the planes (hkl) in a crystal is completely specified if the direction of the plane normal $\mathbf{n}$ is known. Figure 1A shows a hypothetical orientation sphere of unit radius and area equal to 4π. The total number of possible

[4] S. K. Allison, Phys. Rev. **41**, 1 (1932).
[5] R. M. Bozorth and F. E. Haworth, Phys. Rev. **45**, 821 (1934).
[6] P. Kirkpatrick and P. A. Ross, Phys. Rev. **43**, 596 (1933).
[7] B. Davis and W. M. Stempel, Phys. Rev. **17**, 608 (1921).
[8] W. L. Bragg, R. W. James, and C. H. Bosanquet, Phil. Mag. **42**, 1 (1921).

orientations of the crystal is proportional to 4π, but the number of orientations compatible with the Bragg condition is proportional to the area of the zone of width $\Delta\theta_i$ and circumference $2\pi\cos\theta$. Thus, considering the entire cone of diffracted rays, the proportion of favorable orientations is the ratio of the area of this zone to the area of the sphere, or

$$A'/A = (2\pi\Delta\theta_i\cos\theta/4\pi) = \tfrac{1}{2}\Delta\theta_i\cos\theta.$$

Taking into account the multiplicity, m, of the planes (hkl), the proportion of favorable orientations is given by $(m/2)\Delta\theta_i\cos\theta$.

The above development does not take into account the limitations imposed by the geometrical features of the experimental set-up. The width of the zone $\Delta\theta_i$ is increased as a result of the finite width of the x-ray source. On the other hand, only a small segment of the diffraction halo is intercepted by the receiving slit. The actual degree of freedom of orientation permitted can be defined in terms of rotation of the plane normal $\mathbf{n}$ about two perpendicular axes, giving increments $\Delta\alpha$ and $\Delta\beta$ and an element of area $\Delta\alpha\Delta\beta$ on the surface of the orientation sphere. The ideal proportion of favorable orientations for an actual experimental set-up is then

$$p = (m\Delta\alpha\Delta\beta/4\pi). \tag{7}$$

The quantities $\Delta\alpha$ and $\Delta\beta$ can be determined as follows. Figure 2 shows schematically the arrangement of source, sample, and receiving slit and appropriate orientations of three axes of reference. $\Delta\alpha$ and $\Delta\beta$ are conveniently expressed in terms of rotations about axes 3 and 2 respectively. The degree of freedom about axis 3 is most readily expressed. The rectangle representing the flat sample may now be thought of as a magnified picture of one plane of the family (hkl) in a single crystallite oriented so as to fulfill the Bragg condition. It is seen that $\Delta\alpha$ is composed of two terms, one equal to the rocking angle of the crystallite, $\Delta\theta_i$, and the other proportional to the apparent width of the x-ray tube focus, W_F. Hence we have

$$\Delta\alpha = \Delta\alpha_{W_F} + \Delta\theta_i = W_F/L + \Delta\theta_i. \tag{8}$$

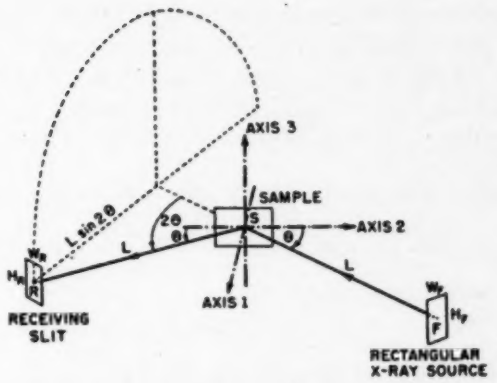


FIG. 2. Certain geometrical aspects of diffraction using the asymmetric focusing arrangement.

$\Delta\beta$ also consists of two terms, $\Delta\beta_{HF}$ and $\Delta\beta_{HR}$, which depend respectively upon the heights of the x-ray tube focus and the receiving slit. Referring to Figs. 1A and 1B, F is the x-ray tube focus, S the flat powder specimen (or it may be regarded as one plane of the family (hkl) in a crystallite oriented so as to fulfill the Bragg reflection condition), R the receiving slit, L the distance SF or SR, and $\mathbf{n}$ the vector normal to the planes (hkl). In Fig. 1B imagine oscillation of a diffracting crystallite about an axis coincident with the diffracted ray, SR. A diffracted ray will continue to be received at R as long as the oscillation does not exceed an angular rotation increment $\Delta\beta_{SR}$, which is governed by the height of the focus (perpendicular to the plane of the paper). This angular degree of freedom is

$$\Delta\beta_{SR} = H_F/L\sin 2\theta.$$

Since we are interested in the degree of freedom about axis 2 rather than axis SR, the angular rotation vector $\boldsymbol{\Delta\beta}_{SR}$ may be resolved into two rectangular components parallel to axes 1 and 2, $\boldsymbol{\Delta\beta}_1$ and $\boldsymbol{\Delta\beta}_2$ respectively (refer to insert, Fig. 1). But $\boldsymbol{\Delta\beta}_1$ is equivalent to rotation about the plane normal $\mathbf{n}$ and has no bearing on the diffraction effects. We are left with the component $\boldsymbol{\Delta\beta}_2$, which is identical with $\boldsymbol{\Delta\beta}_{HF}$. The magnitude of this vector is

$$\Delta\beta_{HF} = \Delta\beta_{SR}\cos\theta = H_F\cos\theta/L\sin 2\theta = H_F\operatorname{cosec}\theta/2L. \tag{9}$$

A similar line of reasoning applies in the evaluation of $\Delta\beta_{HR}$. Referring to Fig. 1A, imagine oscillation about an axis coincident with SF.

The degree of freedom is limited by the height of the receiving slit, H_R, and is seen to be $\Delta\beta_{SF}=H_R/L\sin 2\theta$. Again this rotation must be resolved into components about axes 1 and 2, and the magnitude of the rotation about axis 2 is

$$\Delta\beta_{H_R}=H_R\cos\theta/L\sin 2\theta=H_R\,\mathrm{cosec}\theta/2L. \quad (10)$$

From Eqs. (9) and (10) we then obtain for $\Delta\beta$

$$\begin{aligned}\Delta\beta&=\Delta\beta_{H_F}+\Delta\beta_{H_R}\\&=(1/2L)(H_F+H_R)\mathrm{cosec}\theta. \quad (11)\end{aligned}$$

Substitution of (8) and (11) in (7) gives for the proportion of favorable orientations

$$p=(m/8\pi L^2)(W_F+L\Delta\theta_{\frac{1}{2}})(H_F+H_R)\mathrm{cosec}\theta, \quad (12)$$

in which m is the multiplicity factor of the reflecting planes, L is the sample-to-source and sample-to-receiver distance, W_F is the apparent width of the x-ray tube focus (source), H_F is the height of the focus, H_R is the height of the receiving slit, $\Delta\theta_{\frac{1}{2}}$ is the rocking angle of the crystal for the given reflection, and θ is the Bragg angle.

III. HYPOTHETICAL CASE OF A NON-ABSORBING SAMPLE

This case is never actually realized, but it may be closely approximated by a very thin sample of powder consisting only of materials of low atomic weights. The case is of particular interest because the quantity N in Eqs. (5) and (6), the total number of particles irradiated, has direct physical significance, and the relative intensity deviation can be calculated directly by simply substituting appropriate values for N, p, and q in Eq. (6).

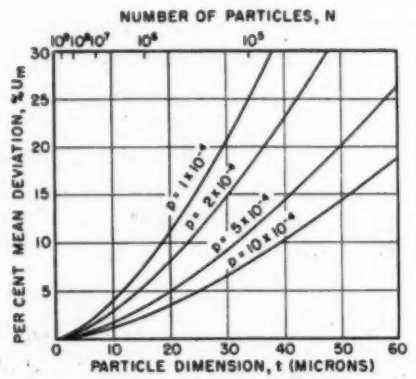


FIG. 3. Theoretical dependence of intensity deviations upon particle dimensions, and number of particles irradiated for the hypothetical case of no absorption.

In the present discussion we are concerned only with the case of the incident x-ray beam impinging upon the surface of a flat powder cake at an angle θ. For a beam of cross-sectional area A the area of sample surface irradiated is then $A/\sin\theta$. If d is the sample thickness, ρ the density of the solid material, and ρ' the density of the powder, the volume of crystalline material irradiated is given by $V=(\rho' Ad/\rho\sin\theta)$. For particles of a uniform volume v the number of particles irradiated is

$$N=V/v=(\rho' Ad/\rho v\sin\theta). \quad (13)$$

In the present non-absorbing case the intensity of the diffracted beam is directly proportional to the number of particles favorably oriented. It follows that for a series of intensity measurements from successive samples of the same powder, the *relative* deviation of a given measurement from the mean will be equal to the *relative* deviation from the mean of the number of particles favorably oriented. The relative mean deviation is given by Eq. (6):

$$U_m=0.798(q/Np)^{\frac{1}{2}}.$$

We can now calculate values of U_m directly by substituting Eqs. (12) and (13) respectively for p and N in Eq. (6). Appropriate values must, of course, be assigned to the beam area A, sample depth d, particle volume v, and the densities ρ and ρ'.

Figure 3 shows the variation of the percent relative mean deviation (% U_m) with particle dimension, t, and number of particles irradiated, N, for the following typical values of the quantities listed above: $A=0.10$ cm^2, $d=0.02$ cm, $\sin\theta=0.25(\theta=14.5°)$, and $\rho'/\rho=0.50$. From Eq. (13) the number of particles is then related to the particle volume, v, and the particle edge dimension, t (assuming cubical particles), by the equation,

$$N=0.004/v=0.004/t^3.$$

A consideration of the various quantities contributing to the value of p for typical experimental conditions (see Section VII) indicates that it is expected to vary between the limits 10^{-3} and 10^{-4} for most cases. For very low angle

reflections it may be somewhat larger, while for an exceptionally small focal spot and short receiving slit it may be smaller. The multiplicity factor of the reflecting planes also has a large effect on the value of p as is evident from Eq. (12), and, hence, very small or very large m's also may cause p to be smaller or larger than the usual range of 10^{-3} to 10^{-4}. In Fig. 3 the four curves portray the results to be expected for representative values of p such as may be expected in actual practice. The curves show clearly that a very small particle size is required for good reproducibility of intensity measurements. Except for unusually favorable values of p the particle dimension must be less than 5 microns for a mean deviation in measured intensity of less than one percent. The equivalent number of particles irradiated must be more than ten million.

IV. CASE OF AN ABSORBING SAMPLE OF UNIFORM PARTICLE SIZE

The x-ray beam of cross-sectional area A impinges upon the surface of the flat powder sample at the angle θ, the correct Bragg angle for diffraction from a set of planes (hkl). The diffracted beam, therefore, also makes an angle θ with the sample surface (Fig. 4). A ray diffracted from a small element of powder at a depth x has traveled a distance s through the powder. The picture is simplified by assuming the sample to consist of only solid crystalline material ($\rho'/\rho=1$ in the notation of Section III). The thickness must be sufficiently great to give complete absorption. It is customary to stipulate as the criterion for this condition that $\mu s \geq 6.4$, μ being the linear absorption coefficient of the solid material if we accept the assumption just made that $\rho'/\rho=1$. In practice the powder contains interstices and ρ'/ρ is less than unity, but this in no way invalidates the discussion to follow.

The area irradiated by the beam is $A/\sin\theta$ as in the case of a non-absorbing sample (Section III). However, in the present case the intensity of the ray diffracted from a particle decreases exponentially with the depth x. Consider, then, the diffracted intensity from a layer of thickness dx at depth x. The volume of this layer is

$$dV=(A/\sin\theta)dx,$$

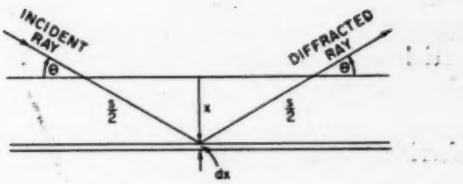


FIG. 4. Diffraction from a layer of powder of thickness dx at depth x.

or, since $x=\frac{1}{2}s\sin\theta$,

$$dV=\tfrac{1}{2}Ads.$$

Let I_0 be the intensity of the ray diffracted at the angle θ by unit volume of properly oriented particles under conditions of non-absorption. The intensity diffracted by one particle of volume v at depth x is then

$$I=vI_0e^{-\mu s}. \tag{14}$$

Now $1/v$ is the number of particles per unit volume and p is the proportion of particles favorably oriented, so that the intensity of the beam diffracted from the volume element dV is

$$\begin{aligned} dI&=(pdV/v)\cdot vI_0e^{-\mu s}\\ &=\tfrac{1}{2}pAI_0e^{-\mu s}ds. \end{aligned}$$

The intensity of the entire diffracted beam is then

$$I=\int_0^\infty dI=I_0pA/2\mu. \tag{15}$$

Equation (15) gives the theoretical intensity to be expected if the actual number of particles favorably oriented is precisely Np. We wish to express the deviation in the actual intensity from this theoretical value as the result of statistical influences. As in Section III the standard, probable, or mean deviation will depend upon the total number of particles irradiated, but in the present situation the function is more complicated as the result of absorption.

The absolute deviation in intensity, σ_I, will in general be related to the deviation in the number of particles favorably oriented as follows:

$$\sigma_I=\begin{pmatrix}\text{deviation in number of}\\ \text{particles favorably oriented}\end{pmatrix}\times\begin{pmatrix}\text{intensity diffracted}\\ \text{per particle}\end{pmatrix}.$$

Considering only the volume element dV at

depth x (absorbing path s) and containing dN particles, the standard deviation can be written

$$d\sigma_I = (dNpq)^{\frac{1}{2}} v I_0 e^{-\mu s}, \tag{16}$$

referring to Eq. (5). But

$$dN = dV/v = A\,ds/2v,$$

so that Eq. (16) becomes

$$d\sigma_I = (\tfrac{1}{2}pqAv\,ds)^{\frac{1}{2}} I_0 e^{-\mu s}. \tag{17}$$

It is now necessary to integrate from $s=0$ to ∞. This must be done as follows. The probable deviations δ_1, δ_2, $\delta_3, \cdots$ of a series of measured quantities are related to the probable deviation Δ of the sum of these quantities by the equation,

$$\Delta^2 = \delta_1^2 + \delta_2^2 + \delta_3^2 + \cdots = \sum \delta_i^2.$$

Therefore, the deviation in the total diffracted intensity will be related to the deviations in intensity from the various volume elements as follows:

$$\sigma_I^2 = \sum d\sigma_I^2 = \int d\sigma_I^2. \tag{18}$$

Hence, by substituting Eq. (17) in (18) we obtain

$$\sigma_I^2 = \tfrac{1}{2} I_0^2 pqAv \int_0^\infty e^{-2\mu s} ds$$

$$= I_0^2 pqAv/4\mu.$$

The standard deviation in the total intensity, therefore, is given by

$$\sigma_I = \frac{I_0}{2}(pqAv/\mu)^{\frac{1}{2}}. \tag{19}$$

The relative standard deviation can be obtained by dividing Eq. (19) by Eq. (15), and the inclusion of the factor 0.798 gives the relative mean deviation. The result is

$$U_m = 0.798\sigma_I/I = \frac{0.798\dfrac{I_0}{2}(pqAv/\mu)^{\frac{1}{2}}}{I_0 pA/2\mu}$$

$$= 0.798(q\mu v/pA)^{\frac{1}{2}}. \tag{20}$$

By comparison of Eqs. (20) and (6) it is seen that the quantity $A/\mu v$ is an effective number of particles, which may be denoted by N_e. Since in general $N = V/v$ [Eq. (13)], this leads to an effective volume of irradiated material, V_e, which has the value A/μ. These quantities, then, are defined by:

Effective Volume Irradiated, $V_e = A/\mu$. (21)

Effective Number of Particles,

$$N_e = V_e/v = A/\mu v. \tag{22}$$

V. CASE OF A PARTICLE SIZE DISTRIBUTION

The powder particles are now assumed to possess discrete sizes with volumes v_1, v_2, $v_3, \cdots$ $v_i, \cdots$, but, otherwise, the assumptions presented in the first paragraph of Section IV apply. Again let I_0 be the intensity of the ray diffracted at the angle θ by unit volume of properly oriented particles under conditions of non-absorption. Suppose also that particles of the ith kind constitute a fraction f_i of the total volume of material irradiated.

In this case the intensity of the beam diffracted from the volume element dV may be expressed as the sum of infinitesimal contributions from each of the sizes present. That is,

$$dI = dI_1 + dI_2 + dI_3 + \cdots + dI_i + \cdots$$
$$= \sum dI_i. \tag{23}$$

The intensity of the ray diffracted by one properly oriented particle of the ith kind at depth x is $I_0 v_i e^{-\mu s}$. The number of particles per unit volume of the ith kind is f_i/v_i. Hence, the intensity diffracted from particles of the ith kind in the volume element dV is

$$dI_i = I_0 v_i e^{-\mu s} p(f_i/v_i) dV$$
$$= I_0 p f_i e^{-\mu s} dV. \tag{24}$$

Equations (23) and (24) give for the contribution of the volume element dV,

$$dI = I_0 p(\sum f_i) e^{-\mu s} dV.$$

Since $\sum f_i = 1$ and $dV = \tfrac{1}{2} A\,ds$,

$$dI = \tfrac{1}{2} I_0 pA e^{-\mu s} ds.$$

Integration from 0 to ∞ gives for the intensity of the entire diffracted beam,

$$I = I_0 pA/2\mu, \tag{25}$$

which is seen to agree with Eq. (15) for particles of uniform size.

Considering first only the volume element dV, the net intensity deviation is related to the contributions of the different constituent sizes as follows:

$$d\sigma_I^2 = \sum d\sigma_{I_i}^2. \quad (26)$$

By analogy to Eq. (16), the standard deviation in intensity due to particles of the ith kind is

$$d\sigma_{I_i} = (dN_i pq)^{\frac{1}{2}} v_i I_0 e^{-\mu s}.$$

Employing Eq. (26), we then get for the squared standard deviation in intensity for the entire element dV

$$d\sigma_I^2 = \sum dN_i pq v_i^2 I_0^2 e^{-2\mu s}.$$

Substitution of $f_i A ds/2v_i$ for dN_i gives

$$d\sigma_I^2 = \tfrac{1}{2} pqA I_0^2 (\sum v_i f_i) e^{-2\mu s} ds. \quad (27)$$

As in Section IV integration from $s=0$ to ∞ must be carried out by employing Eq. (18). This gives

$$\sigma_I^2 = (pqA I_0^2/4\mu) \sum v_i f_i,$$

and we obtain for the standard deviation in the total diffracted intensity

$$\sigma_I = \tfrac{1}{2} I_0 [(pqA/\mu) \sum v_i f_i]^{\frac{1}{2}}. \quad (28)$$

The relative mean deviation is obtained by dividing Eq. (28) by Eq. (25) and multiplying by the factor 0.798. The result is

$$U_m = 0.798 \sigma_I / I = 0.798 \left(\frac{q\mu \sum v_i f_i}{pA} \right)^{\frac{1}{2}}. \quad (29)$$

Comparison with Eq. (20) shows that $\sum v_i f_i$ is the effective particle size, v_e, for a distribution of discrete sizes. We can, then, define the following quantities:

Effective Volume Irradiated, $V_e = A/\mu$. (30)

Effective Number of Particles, $N_e = V_e/v_e = A/\mu v_e$. (31)

Effective Volume per Particle, $v_e = \sum v_i f_i$. (32)

VI. EXPERIMENTAL PROCEDURE AND RESULTS

To test the applicability of the theory, intensity measurements were made on fine quartz powder using a Norelco Geiger-counter x-ray spectrometer. Quartz possesses a number of advantages as the subject for the present study. Single crystals of high purity can be found in nature. Its crystallinity is very perfect, assuring good quality reflections. Quartz is so hard that considerable grinding will not reduce the average particle size sufficiently to produce line broadening. Most important of all, quartz displays only slight preferential cleavage and usually fractures conchoidally, so that the powder crystallites would be expected to display essentially random crystallographic orientations.

Clear colorless quartz crystals were crushed in a Plattner "diamond" mortar until the bulk of the fragments passed a 325-mesh screen. The coarser fraction was discarded. The finer portion was then separated by sedimentation from methanol into four size fractions comprising the approximate ranges 5–50, 15–50, 5–15, and <5 microns. The <5-micron fraction was separated first. Some of the 5–50-micron residue was then saved while the remainder was refractionated into 5–15- and 15–50-micron intervals.

Multiple samples of each size fraction were examined microscopically with the aid of a United States Bureau of Mines type microprojector for dust counting[9] with a magnification of 1000×. A little of the powder was highly dispersed in methanol and then allowed to settle a few minutes in the glass-covered cavity of a special specimen slide. Every particle that appeared in the field of view was then counted and its volume estimated, the process being repeated for many movements of the specimen on the microscope stage. By examining several hundred particles in this way, it is believed that a fairly representative picture of the entire sample was obtained.

Each particle was regarded as a rectangular parallelepiped, a short and long mean dimension being estimated and the third (vertical) dimension being assumed equal to the shorter observed dimension. The particle volume was then taken as equal to the product of these three dimensions. The lower limit of visibility extended to particles of about one micron edge length. Smaller particles undoubtedly were present, but their contribution to the total volume (and therefore to the total intensity) was insignificant. Using formula (32) the effective volumes per particle for each size distribution were calculated with

[9] C. E. Brown, L. A. H. Baum, W. P. Yant, and H. H. Schrenk, U. S. Bureau of Mines Report of Investigations No. 3373, January (1938).

TABLE I. Intensity measurements on fractions of <325-mesh quartz powder. Tabulated values are areas in arbitrary units of the 3.33A maximum as counted with the Geiger-counter x-ray spectrometer using CuKα radiation.

Specimen No.	15–50-micron fraction	5–50-micron fraction	5–15-micron fraction	<5-micron fraction
1	7,612	8,688	10,841	11,055
2	8,373	9,040	11,336	11,040
3	8,255	10,232	11,046	11,386
4	9,333	9,533	11,597	11,212
5	4,823	8,530	11,541	11,460
6	11,123	8,617	11,336	11,260
7	11,051	11,598	11,686	11,241
8	5,773	7,818	11,288	11,428
9	8,527	8,021	11,126	11,406
10	10,255	10,190	10,878	11,444
Mean area:	8,513	9,227	11,268	11,293
Mean deviation:	1,545	929	236	132
Mean % deviation:	18.2	10.1	2.1	1.2

the following results:

15–50 microns, 394×10^{-10} cm³
5–50 microns, 329×10^{-10} cm³
5–15 microns, 32.1×10^{-10} cm³
<5 microns, 0.4×10^{-10} cm³

The intensity of the 3.33A maximum of quartz was measured using CuKα radiation. The maximum in each case was scanned manually and the counts plotted, the area above background being taken as proportional to the intensity. The customary simplification was adopted of setting the area equal to the product of the peak height and the width at half maximum intensity. The effect of Geiger-counter statistical errors upon the measured intensities was kept well below one percent by taking a sufficient number of counts at each point during the manual scanning.

For each size fraction ten different samples of powder were measured under as nearly identical experimental conditions as possible. The diffraction specimens were prepared by pressing the powder samples into the cavities of biological drop slides. The dimensions of these cavities are such that the entire incident beam is received by the powder surface, and the depth is sufficient to yield a diffracted beam of maximum intensity.

TABLE II. Comparison of observed mean intensity deviations with theoretical values.
$p = 0.00386(0.031 + 13\Delta\theta_i)$.
$\%\ U_m = 1614(v_s/p)^{\frac{1}{2}}$.

Size range (microns)	Effective particle volume, v_s ($cm^3 \times 10^{10}$)	Calculated % U_m for various values of $\Delta\theta_i^{\frac{1}{2}}$ 5′	10′	20′	30′	Observed value of % U_m
15–50	394	23.1	19.6	15.8	13.3	18.2
5–50	329	21.1	18.0	14.4	12.2	10.1
5–15	32.1	6.6	5.6	4.5	3.8	2.1
<5	0.40	0.7	0.6	0.5	0.4	1.2

Table I shows the results of the measurements on the four powder fractions, the measured areas being expressed in arbitrary units. The relationship between particle size and the reproducibility of the diffracted intensities is strikingly demonstrated. For the coarsest fraction of powder the mean deviation in intensity from the average of the 10 values is 18.2 percent, while for the finest fraction it is only 1.2 percent. It will also be noticed that the average observed intensity is smaller for the larger particle sizes, an effect which must be attributed to extinction. It is known that quartz exhibits high crystalline perfection,[5] and accordingly it will be expected to display extinction effects down to very small crystallite sizes as in the case of calcite.

VII. COMPARISON OF EXPERIMENT AND THEORY

The observed intensity deviations may now be compared with theoretical values by employing the appropriate equations developed in Section V for the case of a particle size distribution. Equation (29) cannot be used directly because of the circumstance that the 3.33A quartz powder reflection consists of contributions from planes $(10\bar{1}1)$ and $(01\bar{1}1)$, each set having a multiplicity of six.[10] The relative intensities[11] of these two reflections are approximately 2:1.

Consider the general case of two component reflections 1 and 2 with multiplicities m_1 and m_2 and intensities I_1 and I_2. Making use of Eq. (25), the total intensity can be expressed as

$$I = I_1 + I_2 = \tfrac{1}{2}(A/\mu)[p_1(I_0)_1 + p_2(I_0)_2]. \quad (33)$$

The squared standard deviation in the total intensity is given by

$$\sigma_I^2 = (\sigma_{I_1})^2 + (\sigma_{I_2})^2,$$

and, expressing σ_{I_1} and σ_{I_2} by formulae analogous to Eq. (28), we obtain

$$\sigma_I^2 = A/4\mu(\sum v_i f_i)[p_1 q_1 (I_0)_1^2 + p_2 q_2 (I_0)_2^2].$$

Then σ_I is given by

$$\sigma_I = \tfrac{1}{2}(A \sum v_i f_i/\mu)^{\frac{1}{2}}[p_1 q_1 (I_0)_1^2 + p_2 q_2 (I_0)_2^2]^{\frac{1}{2}}. \quad (34)$$

[10] *Internationale Tabellen zur Bestimmung von Kristallstrukturen* (Gebrüder Borntraeger, Berlin, 1935), Vol. 2, p. 502.
[11] R. E. Gibbs, Proc. Roy. Soc. **110A**, 443 (1926).

TABLE III. Theoretical percent mean deviation in diffraction intensities as a function of particle size. $\%\ U_m = 9744(\mu v_e)^{\frac{1}{3}}$; particles assumed to be cubical.

Effective particle dimension, l_e (microns)	Effective particle volume, v_e ($cm^3 \times 10^{12}$)	Calculated $\%\ U_m$ for various values of the linear absorption coefficient, μ				
		5	20	100	500	2000
1	1	0.02	0.04	0.1	0.2	0.4
2	8	0.06	0.1	0.3	0.6	1.2
5	125	0.2	0.5	1.1	2.4	4.9
10	1,000	0.7	1.4	3.1	6.9	13.8
20	8,000	2.0	3.9	8.7	19.5	39.0
30	27,000	3.6	7.2	16.0	35.8	—
40	64,000	5.5	11.0	24.7	—	—
50	125,000	7.7	15.4	34.5	—	—
CuKα radiation:		Organic compounds	Metal-organic compounds	Siliceous minerals	Cu, Ni, TiO_2, $CdSO_4$	Ag, Pb, PbO, HgO

Referring to Eq. (29), the relative mean deviation is then

$$U_m = 0.798\sigma_I/I = 0.798(\mu \sum v_i f_i/A)^{\frac{1}{3}} \times \frac{[p_1 q_1 (I_0)_1^2 + p_2 q_2 (I_0)_2^2]^{\frac{1}{3}}}{[p_1 (I_0)_1 + p_2 (I_0)_2]}. \quad (35)$$

When $m_1 = m_2$, and therefore $p_1 = p_2$ and $q_1 = q_2$ (as in the present instance), Eq. (35) reduces to

$$U_m = 0.798(q\mu \sum v_i f_i/pA)^{\frac{1}{3}} \times \frac{[1+(I_0)_2^2/(I_0)_1^2]^{\frac{1}{3}}}{[1+(I_0)_2/(I_0)_1]}. \quad (36)$$

This is seen to reduce to Eq. (29) when $(I_0)_1$ is set equal to $(I_0)_2$ except that p in (29) is replaced by $2p$ in (36). This is the proper result since the case of two component reflections for which $m_1 = m_2$ and $I_1 = I_2$ is identical with the case of a single reflection for which $m = 2m_1 = 2m_2$ and $p = 2p_1 = 2p_2$.

We may now proceed to substitute the proper experimental values in Eq. (36). For quartz and CuKα radiation $\mu = 92.8$. The cross sectional area, A, of the x-ray beam at the point of intersection with the sample was 0.126 cm^2. As explained at the beginning of the present section we may let $(I_0)_1 = 2(I_0)_2$. Substitution of these quantities in (36) gives

$$\%\ U_m = 1614(v_e/p)^{\frac{1}{3}}, \quad (37)$$

where $v_e = \sum v_i f_i$ for a given size distribution. The determination of the appropriate values of v_e for the four size distributions under consideration has been described in Section VI. It only remains to determine the quantity p.

Equation (12) is an analytical expression for p, the proportion of favorable orientations. The following quantities were known or could be determined with fair accuracy from the known dimensions of the experimental set-up:

$W_F = 0.031$ cm $L = 13$ cm
$H_F = 0.28$ cm $\theta = 13.33°$
$H_R = 0.35$ cm $m = 6$.

This gives for p as a function of the full width at half maximum intensity of the rocking curve of the quartz powder:

$$p = 0.00386(0.031 + 13\Delta\theta_{\frac{1}{2}}).$$

As explained in Section II there is considerable uncertainty in $\Delta\theta_{\frac{1}{2}}$. For average single crystals of rock salt it is found to vary from about 5′ to 30′. Since we can only surmise that a comparable degree of imperfection will prevail in crushed quartz crystallites, it is probably best to calculate p and U_m for a range of plausible values of $\Delta\theta_{\frac{1}{2}}$. Thus for $\Delta\theta_{\frac{1}{2}} = 5$, 10, 20, and 30′ the corresponding values of p are respectively 1.92, 2.66, 4.11, and 5.77×10^{-4}. Table II presents the results of the calculations for each of the particle size distributions being considered. The results are compared with the experimentally observed intensity deviations.

In comparing the theoretical and experimental results it should be borne in mind that the degree of agreement to be anticipated is conditioned by several relatively large sources of error: (a) The measurement of the dimensions of several hundred randomly selected particles has been assumed to lead to a truly representative value of the effective particle volume, v_e, for the entire

sample consisting of a million or more particles. (b) The observed mean intensity deviations are based in each case upon only 10 observations, and the likelihood that such values will closely approximate the true deviations (for an infinite number of observations) is small. (c) The actual rocking angle for the powder crystallites is not known and, hence, the theoretical deviations must be computed on the basis of a range of values of $\Delta\theta_{\frac{1}{2}}$ which is believed to include the true value. (d) The powder crystallites are assumed to be uniformly imperfect throughout, a predication which certainly oversimplifies the picture and may not be justifiable. (e) Extinction effects are known to exist but they have been neglected in the present analysis. (f) The theoretical intensity deviations are calculated on the basis of the simplified Laplacian Eq. (3), whereas Eq. (4) must be employed for good accuracy when the deviations are much over 10 percent. The effect is to make these calculated deviations somewhat too large.

In view of the considerable sources of error just enumerated the degree of agreement displayed by the data of Table II is certainly rather satisfactory. It demonstrates that the broader aspects of the theory are correct, even though a completely satisfactory expression of the many details has not been accomplished.

VIII. DISCUSSION

The present experimental results for quartz powder demonstrate the great importance of particle size as related to the reproducibility of intensity measurements for reflections from stationary samples. In order to make the picture more complete the data in Table III have been computed by making use of Eq. (29). This table presents the theoretical percent mean deviation in measured intensities as a function of the effective particle dimension, t_e, and volume, v_e, for a wide range of absorbing powers. Typical values for the various geometrical dimensions of the experimental set-up have been used. Thus, W_F, H_F, H_R, L, A, and θ have been assigned the values which are given in Section VII. A multiplicity factor of $m=12$ and a rocking angle of $\Delta\theta_{\frac{1}{2}}=10'$ are assumed, the latter value being appropriate for moderately imperfect crystals. Under these conditions p, the proportion of favorably oriented particles, is equal to 5.32×10^{-4}, a figure which is typical for ordinary experimental conditions. When these values are substituted in Eq. (29) the result is

$$\%\ U_m=9744(\mu v_e)^{\frac{1}{2}}.$$

The validity of this result is dependent on the sample's being of sufficient thickness to give a diffracted ray of maximum intensity. In order to facilitate the interpretation of the data presented in Table III, typical materials of each absorbing class are appended for the particular case of CuKα radiation.

Table III can serve as an approximate guide for experimental work in which a given level of precision is desired. For example, suppose a reproducibility of ±1 percent is aimed at. Assuming that instrumental and human errors can be reduced below this level, the desired precision can be hoped for only if the particle size of organic compounds is reduced below about 10 microns, that of siliceous minerals below 5 microns, and that of PbO below 2 microns. On the other hand, if instrumental and human errors amount to ±2 percent, it is useless to seek to improve the reproducibility by reducing the sizes of the above materials below about 20, 8, and 3 microns respectively.

In this connection it should be pointed out that the mean observed intensity deviation for <5-micron quartz powder (Table II) is larger than the theoretical values because of a minimum precision limit of about one percent which is imposed by the residual instrumental and human errors. This is borne out by the fact that with decreasing particle size the observed deviations decrease more rapidly than the calculated values through the first three size fractions listed, whereas the observed value becomes inconsistent with this trend for the smallest size fraction.

This more rapid falling-off of the observed deviations with decreasing particle size is the most conspicuous point of disagreement between theory and experiment. Further investigation has shown that it arises, in part at least, from the probably unwarranted assumption that the interior of a crystallite is uniformly imperfect throughout (uniform values of $\Delta\theta_{\frac{1}{2}}$ and p). Bozorth and Haworth[5] and subsequently Arm-

strong[12] have shown that the degree of imperfection induced in quartz crystals by abrasive treatments is a function of depth below the surface. Armstrong's work indicates a surface layer approximately ½ micron thick which consists of crystalline material misoriented by a relatively large amount (45′ to 4½°), while a lesser degree of imperfection (misorientations from 0′ to 45′) persists to depths of 50 to 100 microns.

It is reasonable to expect that a somewhat similar condition exists within the particles produced by crushing and grinding quartz and other rather perfect crystals. The present statistical theory has been extended to cover the case of particles consisting of an imperfect surface layer and a relatively perfect interior. Unfortunately this development cannot be carried to completion without a determination of the proper extinction coefficient to characterize the crystal interiors. The experimental solution of this problem is difficult and perhaps impossible. However, preliminary calculations show that there probably exists an optimum empirical value of the extinction coefficient which leads to greatly improved agreement between theory and experiment over that presented in Table II. That is, it appears that the proper incorporation of extinction into this statistical theory will cause the calculated intensity deviations to fall off more rapidly with decreasing particle size than those derived on the basis of the simpler theory. It is hoped that a satisfactory analysis of this kind can be presented at a later date.

[12] E. J. Armstrong, Bell Sys. Tech. J. **25**, 136 (1946).

IX. ACKNOWLEDGMENTS

The authors are indebted to Dr. J. T. Kummer, Mellon Institute, and Dr. E. W. Montroll, University of Pittsburgh, for helpful discussions during the development of the theoretical portions of this paper. They also wish to thank members of the Industrial Hygiene Foundation staff for valuable help in the sedimentation and microscopic examination of the powder samples.

Methods for Measuring Dynamic Mechanical Properties of Rubber-Like Materials*

A. W. NOLLE

Acoustics Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received February 3, 1948)

Five methods are described for measuring the complex differential dynamic Young's modulus of rubber-like materials, under conditions of very small sinusoidal strain variations, at frequencies from 10^{-1} to 10^{5} cycles per second and in a temperature range which at its widest limits extends from minus 60 to plus 100 degrees centigrade. These methods are (1) measurement of the frequency of oscillation and the decrement of a "rocking-beam oscillator" in which the restoring force is supplied by rubber (0.1 to 25 c.p.s.); (2) measurement of the frequency of resonance and the band width of a rubber reed which executes forced oscillation (10 to 500 c.p.s.); (3) measurement of the phase velocity and the attenuation of longitudinal waves in a thin rubber strip in the absence of standing waves (1 to 40 Kc); (4) measurement of the frequency of longitudinal-wave resonance and the bandwidth of a half-wave transmission line (below 300 c.p.s.); (5) measurement of the change in the frequency of resonance and the change in the band width of a magnetostrictive rod against whose end a rubber sample is pressed (12 to 120 Kc). Equations are developed for calculating the complex modulus from the results of each experiment, and the practical advantages and limitations of the methods are described.

LIST OF SYMBOLS

Except where specific indication is given to the contrary, c.g.s. units have been used throughout this paper. The convention that a sinusoidal variation with time is represented by the real part of $\exp(-i\omega t)$ has been followed consistently; thus $-i$ as used in this paper is equivalent to the j of electrical engineering. The following list includes all symbols except some which are used to a very limited extent and are defined where used.

* Based on material submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology. This work was supported in part under Contract NObs-25391, Task No. 1, with the Bureau of Ships. A detailed presentation of experimental results obtained by the methods described in this paper will appear in the Journal of Polymer Science.

Symbol	Significance
A	Cross-sectional area of rubber sample, cm².
c	Acoustic phase velocity, cm/sec.
D	$(\nu_e-\nu_0)/\nu_0$, magnetostriction method.
$E^*=E_1-iE_2$	Complex Young's modulus, dynes/cm².
f	Frequency, kc.
F	Force or force amplitude, dynes.
k	Mechanical stiffness, dynes/cm.
l	Static length of a rubber sample, cm. Length of magnetostriction rod, cm.
L	Acoustic attenuation, db/cm.
m	Mass, g.
Q	Relative sharpness of resonance $\nu/\Delta\nu$. (Also used for $m\omega/R$, approximately equal to $\nu/\Delta\nu$.)
r	$(1/2\pi)$ times attenuation in nepers per wave-length.
R	Mechanical resistance, dynes-sec./cm.
t	Thickness of rubber sample, cm. Time, sec.
T	Period of oscillation, etc.
α	Acoustic attenuation, nepers/cm.
α', β	Loss and phase parameters of hyperbolic-tangent impedance method.
γ	Extensional viscosity coefficient (also known as "normal" or "Young's modulus" viscosity coefficient), dynes-sec./cm².
λ	Wave-length, cm.
ν	Frequency, c.p.s.
ξ	Displacement, cm.
ρ	Density, g/cm³.
ω	"Natural" frequency, radians/sec.

I. INTRODUCTION

THE study of the elastic behavior of rubber-like materials under quasi-static conditions has received attention for more than a hundred years. Theories have been developed which explain some aspects of the elastic non-linearity of rubber and which explain the temperature dependence of the static elastic modulus.[1] A smaller amount of study, however, has been devoted to the behavior of rubber-like materials in response to time-dependent forces. Most, but not all, of the studies which have been devoted to this aspect of rubber-like behavior have represented limited technological investigations rather than attempts to make a basic examination of dynamic rubber-like behavior in the simplest terms, with due regard for the influence of all of the pertinent variables.

The response of rubber-like materials to time-variable, or dynamic forces has both theoretical and technical interest. A knowledge of the dynamic behavior of rubber must be added to a knowledge of the static behavior to establish a complete set of phenomena for which the complete theory of rubber-like elasticity must account. The results of a study of the dynamic elastic modulus as a function of frequency and of temperature, for example, give the magnitudes of the molecular energy barriers which must be overcome as the material undergoes a deformation in response to an external force. From the technical point of view the dynamic behavior of rubber-like materials is important because of the usefulness of these materials in absorbing mechanical energy. Rubber-like materials are used to provide low frequency damping in mechanical shock-absorber systems, to provide damping in such audiofrequency devices as phonograph recording elements, and to absorb energy in systems used for the study of ultrasonic propagation in metals.[2] In other technical problems, such as the problem of tire design, a knowledge of dynamic rubber-like behavior is essential to guard against excessive mechanical losses which may lead to excessive heat generation. Because of their relatively large energy-absorbing ability, or viscosity, the rubber-like materials are classed as "viscoelastic" materials. The study of the viscous behavior of the materials is of the same importance as the study of their ordinary elastic modulus in examining their dynamic behavior.

The present research constitutes an attempt to study the dynamic mechanical properties of rubber-like materials over a range of frequency and temperature sufficiently great to encompass the significant variation of the dynamic properties with these variables. In this paper are reported five experimental methods which have been developed in connection with this research. These methods are used to measure the complex differential dynamic Young's modulus at frequencies ranging from 10^{-1} to 10^5 cycles per second. The extremes of the temperature range which has been covered with these various methods are minus 60 and plus 100 degrees centi-

[1] See for example E. Guth, H. M. James, and H. Mark, *Advances in Colloid Science, Vol. II (Rubber)* (Interscience Publishers, Inc., New York, 1946), pp. 253–298.

[2] F. A. Firestone and J. R. Frederick, J. Acous, Soc. Am. 18, 200 (1946).

grade. For simplicity, the measurements are performed with forces and deformations which are sufficiently small that non-linear mechanical response is not encountered.

A considerable amount of literature has been devoted to the general topic of measurements of the dynamic properties of rubber; a smaller portion of this literature is directly concerned with the small-amplitude, sinusoidal, viscoelastic measurements which are of special concern in the present paper.[3] While the methods reported in the present paper are considerably different in detail from methods previously reported, many points of basic similarity will be evident to one familiar with the literature. A number of the earlier methods make use of the principles which apply to four of the methods in this paper—the frequency of resonance of a suitable mechanical system is directly related to the modulus of an incorporated rubber sample, while the bandwidth or the decrement is directly related to the viscous resistance offered by the rubber sample.

In particular, the value of the work of Sack and his associates[3] should be stressed. These investigators, using a vibrating-reed method more elaborate than the one described in the present paper and considerably different in detail, have made a much more extensive study of the effects of temperature and frequency upon the dynamic modulus than is to be found in other existing papers, and have demonstrated the existence of a maximum in the loss-per-cycle *vs.* frequency curves for certain butadiene-styrene copolymers at low temperatures, in the audiofrequency range. The previous investigations of dynamic rubber-like elasticity, including that of Sack and his associates, have been but little concerned with the variation with frequency of the real part of the dynamic modulus. Instead, the frictional loss in the dynamic deformation of rubber has generally been the quantity whose dependence on frequency has received attention. The previously reported investigations have been restricted to frequencies below a few hundred cycles per second, with the exception that the work of Sack extended to frequencies as high as 4000 c.p.s.

The strip transmission method reported in the present paper, an experiment which does not depend upon resonance, is an extension of the Ballou and Silverman modification of the Melde experiment. The Ballou and Silverman experiment[3] is described as a means for finding the Young's modulus of materials, such as Nylon, which have relatively small internal friction, by a sound velocity measurement. The internal friction is not measured.

Definitions of the Experimental Quantities

Modulus will be taken to mean Young's modulus. This quantity will always be stated in dynes/cm^2 and will be denoted by E_1. Specifically, the force contribution due to the Young's modulus when a column of rubber is subjected to a small static elongation is

$$F = E_1 A \xi / l, \tag{1}$$

where l is the natural length of the column along the axis of the force, A is the cross-sectional area, and ξ is the elongation of the column. The force is that due to the purely elastic portion of the deformation; viscous forces are excluded.

It is usually understood that the Young's modulus for rubber is three times the shear modulus, because the high incompressibility of the material requires that Poisson's ratio be $\frac{1}{2}$. More exactly, this relation is valid for each rubber-like material through some range of temperature and frequency for which the modulus does not become excessively large. Under these conditions the three-to-one relation also holds for the viscosity coefficient. The results of dynamic meas-

[3] An account of the earlier work in this field has been given by W. W. Vogt in *The Chemistry and Technology of Rubber* (Reinhold Publishing Company, New York, 1937), pp. 332–379. The more pertinent examples of later work are discussed in the following references: J. W. Ballou and S. Silverman, J. Acous. Soc. Am. **16**, 113 (1944); J. H. Dillon, I. B. Prettyman, and G. L. Hall, J. App. Phys. **15**, 309 (1944); J. H. Dillon and S. D. Gehman, Ind. Rubber World **115**, 16 and 217 (1946); J. D. Ferry, Rev. Sci. Inst. **12**, 79 (1941); W. P. Fletcher and J. R. Schofield, J. Sci. Inst. **21**, 193 (1944); W. P. Fletcher, Nature **153**, 341 (1944); also Rubber Chem. Tech. **17**, 619 (1944); S. D. Gehman, D. E. Woodford, and R. B. Stambaugh, Ind. Eng. Chem. **33**, 309 (1944); C. W. Kosten and C. Zwikker, Physica **4**, 221 (1937); also Proc. Rubber Tech. Conf. London (1938), 987; W. Kuhn, O. Kunzle, and A. Preissman, Helv. Chim. Acta **30**, 307 (1947); M. Mooney and R. H. Gerke, Rubber Chem. Tech. **14**, 35 (1941); J. E. Moyal and W. P. Fletcher, J. Sci. Inst. **22**, 167 (1945), also Rubber Chem. Tech. **19**, 163 (1946); W. J. S. Naunton and J. R. S. Waring, Proc. Rubber Tech. Conf. London (1938) 805; also Trans. Inst. Rubber Ind. **14**, 340 (1939); H. Roelig, Rubber Chem. Tech. **12**, 394 (1939); also Gummi-Ztg.-Kautschuk, June-July 1943, 47–9*K*; also Rubber Chem. Tech. **18**, 62 (1945); H. S. Sack, H. L. Laub, and R. N. Work, J. App. Phys. **18**, 450 (1947); L. B. Sebrell and R. P. Dinsmore, S. A. E. J. **49**, 386 (1941).

urements show, however, that at sufficiently low temperature or at sufficiently high frequency the modulus for distortion approaches the magnitude of the bulk modulus, and the three-to-one relation between Young's modulus and the shear modulus is invalid. Under conditions where Young's modulus and shear modulus measurements are not equivalent, the results obtained by the methods described in this paper apply only to Young's modulus.

Because of the non-linear character of the complete stress-strain curve for rubber, the Young's modulus for rubber is often defined as the ratio of the total stress to the total strain at some arbitrarily specified elongation. This definition of Young's modulus is not suitable for the methods of measurement reported in this paper, because in these methods the variation of strain is sufficiently small that non-linear effects are usually negligible.

Dynamic Modulus refers, in general, to modulus as measured under conditions of time-varying stress and strain. In this report dynamic modulus will refer to modulus measured under conditions of sinusoidally varying stress and strain. Because of the large variation with frequency of the dynamic modulus of rubber, this specification is necessary to give definite significance to the data. The deformations involved in the measurements here described are sufficiently small that the effects of non-linear terms in powers of strain and strain rate higher than the first will not be detectable. Wherever appreciable viscous loss accompanies the deformation, the dynamic modulus will be regarded as a complex quantity in keeping with the definition in the following paragraph. This practice is at variance with the convention of Alfrey and Doty[4] in which dynamic modulus refers to the absolute value of the complex modulus.

Complex modulus refers to a notation which will be used as a means for expressing both the ordinary dynamic *elastic* modulus and the mechanical loss in the material as a complex number. The force associated with resistive loss in the dynamic deformation of a column of rubber may be expressed as

$$F=(\gamma A/l)\partial\xi/\partial t. \tag{2}$$

The quantity γ is a viscosity coefficient, sometimes called the "normal viscosity coefficient," whose geometrical definition agrees with that of Young's modulus rather than that of the usual shear viscosity coefficient. In practice, the real modulus and the viscosity coefficient are functions of frequency. Therefore Eq. (2) is valid only to define an effective value of the viscosity coefficient for constant frequency. Thus it is preferable to relate the definition of the viscosity coefficient to a sinusoidally varying displacement $\xi=\xi_0 e^{-i\omega t}$. Then Eq. (2) is replaced by

$$F=-i\omega\gamma\xi A/l. \tag{3}$$

The total force associated with deformation of the sample, arising from the combination of elastic and viscous effects is

$$F=(A\xi/l)(E_1-i\gamma\omega), \tag{4}$$

provided the dimensions of the sample are small compared to an acoustic wave-length. But this expression is equivalent to the expression for the force arising from a purely elastic deformation, $F=E_1A\xi/l$, if one replaces the ordinary modulus E_1 by the complex modulus E^*. Hence the effective complex modulus at a given frequency is

$$E^*=E_1-iE_2, \tag{5}$$

where

$$E_2=\gamma\omega. \tag{6}$$

It appears from (4) that the complex modulus is related to the mechanical impedance of a unit sample, $Z_m=iF/\omega\xi$, by

$$E^*=-i\omega Z_m, \tag{7}$$

where the impedance Z_m is the mechanical impedance of one face of a unit cube, the opposite face being rigidly backed. When in a practical experiment the complex modulus is to be found from a mechanical impedance measurement by means of Eq. (7), it is necessary that the measurement be performed on a sample small compared to an acoustic wave-length and the result converted to that for a unit cube. An additional important relation is

$$E_2/E_1=\cot\theta, \tag{8}$$

where θ is the phase angle of the mechanical impedance.

Differential dynamic modulus refers to dynamic

[4] T. Alfrey and P. Doty, J. App. Phys. **16**, 700 (1945).

modulus determined from a sample upon which a static deformation is imposed along the same coordinate as the dynamic deformation. Because the differential dynamic modulus is a complicated function of static strain for many rubbers, it it best to regard the sample at any particular static deformation as a distinct material from the standpoint of dynamic modulus measurements. Therefore, the differential dynamic modulus will always be reported with respect to the strained length and the strained cross section of the sample. The differential dynamic modulus cannot be identified with the slope of the quasi-static stress-strain curve except at very low frequency. This statement expresses a special case of the more general observation that the dynamic modulus for all lightly vulcanized rubbers is vastly different from the static modulus, except at frequencies in the very low audio or subsonic frequency ranges.

Loss factor (or "loss tangent") is taken to mean the dimensionless ratio E_2/E_1. This nomenclature is in keeping with the practice of some investigators who define the loss factor of a dielectric as the ratio of the imaginary part of the dielectric constant to the real part. Some confusion exists in this regard because the dielectric "loss factor" is also frequently taken to mean the imaginary part of the dielectric constant.

The following sections describe five experimental methods for the measurement of the differential dynamic Young's modulus which have been used to cover the frequency range from 0.1 c.p.s. to 120 kc. Sample data for the various methods will not be given in this paper, but will instead be incorporated in a more general discussion of the results which will appear under the title "Dynamic mechanical properties of rubber-like materials."

II. THE ROCKING-BEAM OSCILLATOR

The rocking-beam oscillator is a mechanical resonant system in which the restoring force to maintain oscillation is provided by a rubber sample. The methods for finding the real modulus and the viscosity of the rubber sample from the experimentally determined frequency of resonance and damping of a simple mechanical system of this nature are well known. The rocking-beam oscillator was designed to provide a combination of operating characteristics which differs somewhat from that of any of the previous devices in this category, in order to insure that data obtained with the rocking-beam oscillator will be comparable to data obtained with the strip transmission apparatus or with the longitudinal-wave-resonance apparatus (Sections IV and V). The important characteristics of the rocking-beam oscillator are the following:

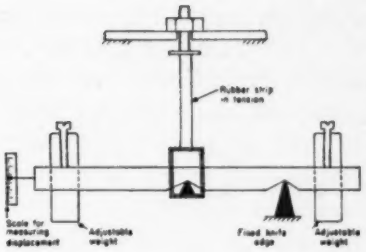


FIG. 1. Sketch of rocking-beam oscillator.

(1) The deformation of the rubber sample is substantially a pure elongation.

(2) An arbitrary static stress can be imposed upon the sample.

(3) The differential elongation of the rubber sample is the *only* effect which produces restoring force for the oscillations.

(4) Experimental observations are made for oscillations corresponding to a differential strain amplitude not greater than 1 to 2 percent; consequently, non-linearity and amplitude-dependence of the frequency, are avoided.

Figure 1 illustrates the rocking-beam oscillator. The rocking beam, when assembled with its associated parts, has its center of gravity at the same height as the knife edge which supports the beam. Consequently, the restoring force for small displacements is due entirely to differential elongation of the rubber sample. The second knife edge serves to connect the rubber sample to the beam in a manner which permits only elongational distortion of the rubber to occur. The period is determined by the moment of inertia of the system (which is controllable through adjustment of the position of the weights on the beam) and by the stiffness of the rubber sample (which is controllable through choice of the dimensions of the sample). A typical sample is 0.1 cm thick, 0.6 cm wide, and from 1.5 to 5 cm long. The static force on the sample can be adjusted by varying the balance of the weights about the central knife edge. This procedure has

little effect on the period. The entire apparatus is located in a temperature control box.

The simplest experimental procedure is to observe the free oscillation of the system visually. Oscillation may be initiated by imparting an initial displacement to the beam by any convenient means. A low power microscope is used to observe the motion in order that a relatively small amplitude of oscillation may be employed. The period of oscillation is measured by means of a manually operated high speed stopclock. The fractional reduction of amplitude over some convenient number of swings is also observed, as a measure of the damping of the system. This method of measurement has been used successfully for frequencies ranging from 0.1 to 3 c.p.s.

An alternative experimental procedure for somewhat higher frequencies makes use of forced vibration of the system. A "pancake" coil is fastened to one end of the beam. The coil moves in the magnetic field between the poles of a small, fixed horseshoe magnet. The system is driven by exciting the coil with the output of a low frequency oscillator. The frequency for maximum amplitude of oscillation is determined, as well as the band width.[5] This modification has been employed at frequencies from 3 to 25 c.p.s. For the higher frequencies in this range the beam shown in Fig. 1 is replaced by a short beam which does not extend beyond the supporting knife edge to the right. A single weight, located very near the support, is used. As an alternative, the movable mass on the beam is eliminated altogether, and the static force necessary to stretch the sample to the desired elongation is applied through a very long rubber band which extends downward from the beam and through a hole in the bottom of the temperature control box. The length of this rubber band is sufficiently great that its dynamic stiffness is negligible compared to that of the sample which is being studied, but the length of the band is not great enough to permit the existence of standing-wave effects in the band.

The apparatus is best suited to critical study of the temperature dependence of the differential dynamic Young's modulus at essentially constant frequency. The uncertainty of the measurements at a given frequency is generally not greater than 1 to 2 percent, but the uncertainty in the actual values of the modulus is somewhat larger than this because of difficulty in securing exact knowledge of the effective length of the sample. There are certainly "end effects" at the clamps in which the sample is terminated, and experimental evidence indicates that the end effects are not completely reproducible. After a series of tests at a given frequency, the rubber sample must be removed in order that the apparatus can be calibrated with a standard spring. Moreover, if the temperature of measurement is carried much above 50°C, the sample suffers a long-period set and must be replaced by a new sample before measurements comparable to the original group can be obtained. Since these considerations require that the sample be replaced after each run at a particular (nearly) constant frequency, it necessarily follows that the data which are obtained for the modulus as a function of temperature at constant frequency are more precise than those which are obtained for the modulus as a function of frequency.

During the course of measurements at various temperatures, the frequency of oscillation varies slightly according to the effect of temperature upon the modulus. It is this slight change in frequency which enables one to calculate the variation of the modulus as a function of temperature. But the frequency variation is only a few percent over the temperature range in which useful measurements are possible—results cannot be obtained when the temperature is so far reduced that the modulus and the damping increase steeply with decreasing temperature, because excessive damping is encountered under these conditions.[6] Since the frequency sensitivity of the

[5] The band width measured in this experiment is the difference between the two frequencies for which the vibration amplitude is $1/\sqrt{2}$ times the maximum amplitude. In the case of a sharp resonance this band width is very nearly equal to the "half-power band width" encountered in electrical engineering, which is the difference between those frequencies for which the square of the current has half its maximum value. Electrical current may be regarded as analagous to mechanical velocity, which is equal in magnitude to radian frequency times displacement.

[6] The rocking-beam oscillator is not suited for the study of a material for which $E_2/E_1 > 0.2$. For larger values of this loss factor, the damping is inconveniently large. This restriction sets a low temperature limit to the measurements for a given material at a particular frequency, because the loss factor for any rubber becomes large when the temperature is reduced sufficiently. While the loss factor again be-

modulus at low frequencies and moderate temperatures is much less pronounced than the temperature sensitivity, the experimental results may for all practical purposes be regarded as modulus-temperature data at constant frequency.

For purposes of analysis the apparatus may be represented as an oscillating system of one degree of freedom, in which a mass, a resistance, and a spring are in series from the electrical point of view. The frequency of free vibration of the system is

$$\nu=(1/2\pi)[k/m-(R/2m)^2]^{\frac{1}{2}}, \tag{9}$$

from which

$$k=m[4\pi^2\nu^2+(R/2m)^2]. \tag{10}$$

The amplitude y diminishes with time according to

$$y=y_0\exp(-Rt/2m). \tag{11}$$

Thus

$$R=(2m/T_n)\log_e(y_0/y_n), \tag{12}$$

where T_n is the time required for n vibrations during which the amplitude diminishes in the ratio y_n/y_0. The real and imaginary parts of the modulus E_1 and E_2 are related to k and R by

$$R=\gamma l/A=E_2 l/\omega A, \tag{13}$$

$$k=E_1 A/l, \tag{14}$$

where A is the cross-sectional area of the sample in cm², and l is the length of the sample in cm. Since the dynamic modulus is a function of static deformation, and the material is therefore effectively a new substance from the dynamic standpoint for each different degree of static deformation, the dynamic modulus will be referred to the deformed cross section of the sample. (This is done automatically in the case of the longitudinal-wave methods of Section IV and V.) Thus l is understood to refer to the *stretched* length of the sample, and A is the associated cross section. Since the volume of the sample is essentially constant for deformations in shape,

$$A=A_0/(1+\epsilon). \tag{15}$$

comes small with a certain additional reduction of temperature, the modulus is in this case increased by a factor of several hundred above its room-temperature value and the system will not oscillate at a useful rate. In practice, the rocking-beam oscillator technique is applicable to a particular rubber only at temperatures 20 or more centigrade degrees above the "static" stiffening temperature.

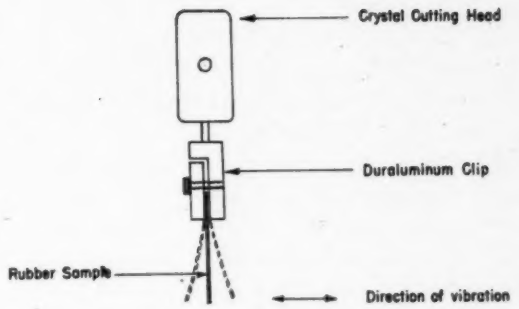


FIG. 2. Sketch of vibrating-reed apparatus. The lateral vibration of the tip of the reed in forced oscillation is observed through a low power microscope.

Here A_0 is the cross section of the undistorted sample, and ϵ is the fractional elongation—for example, $\epsilon=0.5$ when the sample is extended to 150 percent of its natural length.

The effective dynamic mass of the beam could in principle be obtained by computing the moment of inertia of the system, but it is more convenient to calibrate the system for a particular adjustment by substituting a spring for the rubber sample. Suppose that the spring has a period of T_0 seconds with some known mass m_0, and that the rocking beam has a period of T_1 seconds when the spring is substituted for the rubber. Then the effective mass of the rocking beam is

$$m=m_0(T_1/T_0)^2, \tag{16}$$

while E_1 and E_2 are given by

$$E_1=(lm/A)\{(2\pi\nu)^2+(1/T_n)^2\times[(\log_e(y_0/y_n)]^2\}, \tag{17}$$

$$E_2=(4\pi\nu lm/AT_n)\log_e(y_0/y_n). \tag{18}$$

The second term of (17) is not ordinarily required in practice. The damping of the vibration usually prevents attainment of uncertainty much less than 1 percent in the measurement of T^2 (or ν^2). For an actual rubber-like material there is only a very small temperature range in which the low frequency damping is great enough that the magnitude of the second term in (17) approaches 1 percent of the result, but is not so great as to make determination of the period of vibration impracticable.

For the case of forced oscillation the system may be analyzed as a simple series resonant circuit (from the electrical viewpoint) driven by a

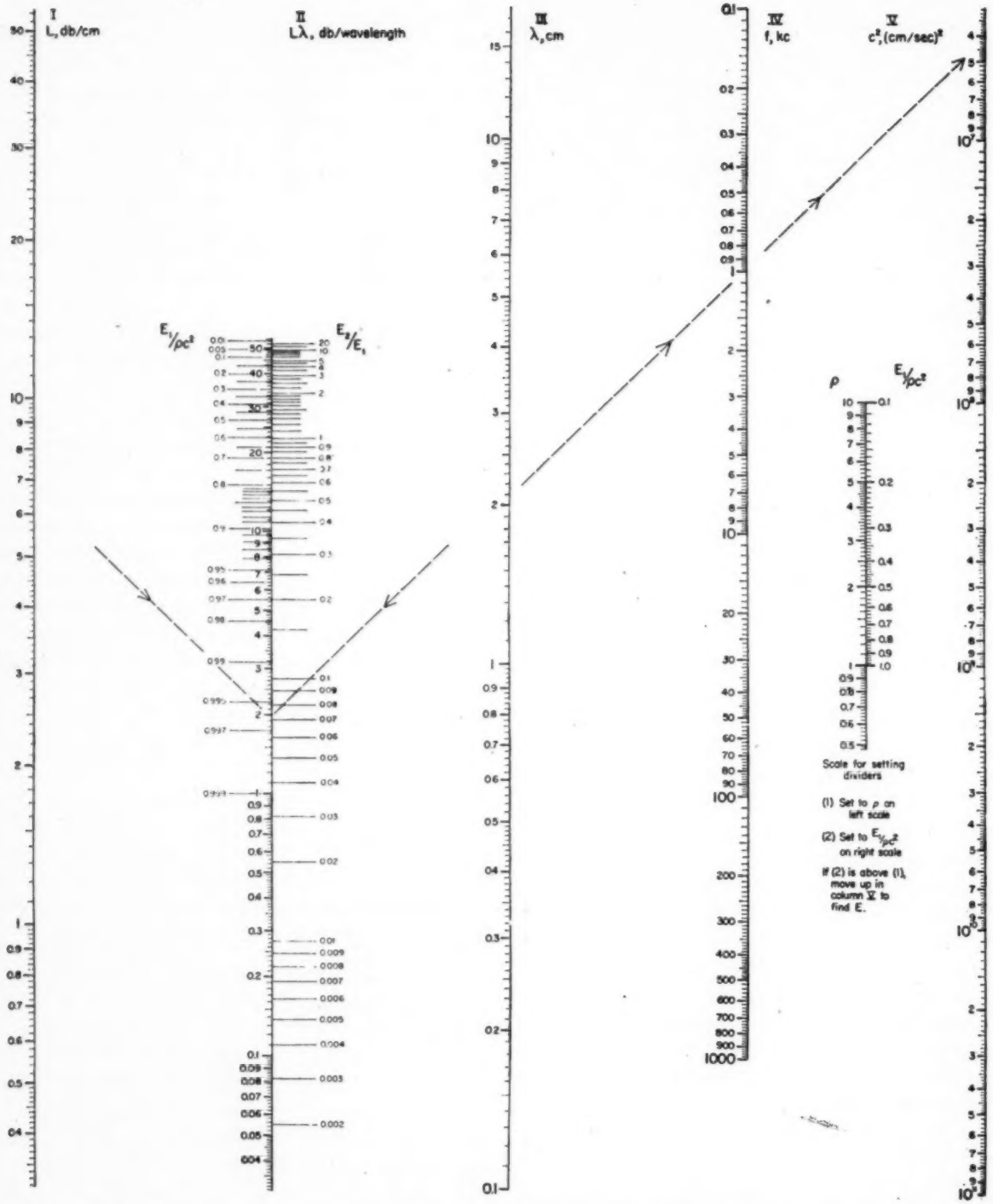


FIG. 3. Logarithmic nomograph for computing the real and imaginary parts of Young's modulus, E_1 and E_2, from the longitudinal wave-length λ and the attenuation in db/cm L measured at frequency f (Kc).

force whose amplitude is independent of frequency. The motion is observed in terms of amplitude (which may be compared to current divided by frequency) rather than in terms of velocity (which may be compared to current). The observed quantities are ν_0, the frequency of maximum amplitude response, and $\Delta\nu$, the band width or difference between those frequencies for which the amplitude has $1/\sqrt{2}$ times its maximum value. The results of the analysis give the approximate relations

$$E_1=(4\pi^2 lm/A)[\nu_0^2+(\Delta\nu)^2/2], \quad (19)$$

$$E_2=4\pi^2\nu_0\Delta\nu lm/A. \quad (20)$$

(Forced oscillation; error <2 percent for $\Delta\nu/\nu_0(\doteq E_2/E_1)<0.2$.)

The quantity m is again obtained by observing the rate of oscillation of the system when a spring is substituted for the rubber sample, and applying (16). The comments given in Appendix I concerning the approximations involved in the equations for the vibrating-reed experiment also apply to the approximations in the derivation of (19) and (20).

III. THE VIBRATING REED

The vibrating-reed technique involves an experimental procedure which is quite similar to that for the rocking-beam oscillator. The resonance phenomenon which is studied in the vibrating-reed experiment is, nevertheless, a standing-wave phenomenon (the flexural-wave resonance of an elastic cantilever). The experimental procedure consists of observing the frequency of resonance and the band width for the forced vibration of a rubber reed. The method is closely related in principle to a vibrating-reed method described by Sack, Laub, and Work.[3] These investigators used electrical means to excite either flexural or longitudinal vibration of a reed with the help of a small driving coil (or in one instance, a thin foil) attached physically to the reed. The present experiment differs from their work in the respect that vibration of the reed is excited by imparting a sinusoidal motion to the reed clamp.

The apparatus is illustrated in Fig. 2. The driving element is a phonograph recording head, designed for lateral motion. Both the piezoelectric and the magnetic types have been used, but it has been found that the magnetic type of unit (provided it is one which does not include internal damping material) is preferable because of its relative constancy of efficiency and of resonance frequency over a wide temperature range. A small aluminum clamp, which holds the end of the rubber cantilever, is terminated in a shank which fits into the recording head in place of the ordinary stylus. The clamp and its shank must have very small mass, in order to place the main mechanical resonance of the loaded cutter at the highest possible frequency. The amplitude of the driver is essentially constant below this resonance, but falls with the inverse second power of frequency for higher frequencies. Consequently, it is impossible to make satisfactory experimental observations of amplitude at frequencies very far above the cutter resonance, and it is therefore desirable to keep the frequency of resonance as high as possible.

In practice it has been possible to place the resonance of the driver, with the reed clamp, at about 600 c.p.s. Experimental observations may be made at frequencies ranging from about 10 to 500 c.p.s. when the driving element is excited by an audio oscillator capable of providing about one watt of power. Vibration at frequencies above 100 c.p.s. can be conveniently observed only for materials which have a "Q" of the order of at least 5 (that is, $E_2/E_1\leqslant 0.2$) although a rough

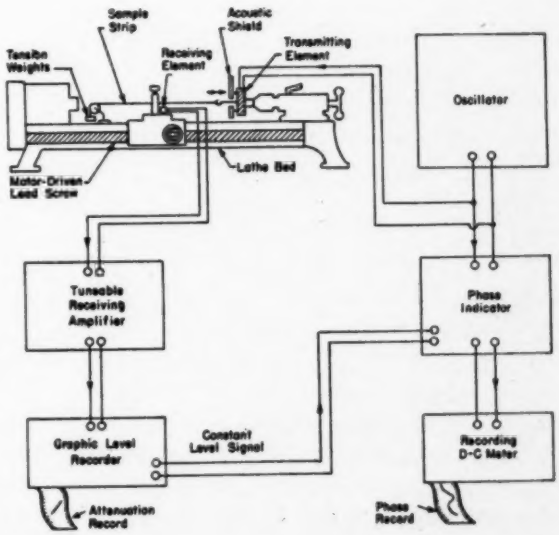


FIG. 4. Schematic diagram of apparatus for measuring longitudinal-wave transmission in thin rubber strips.

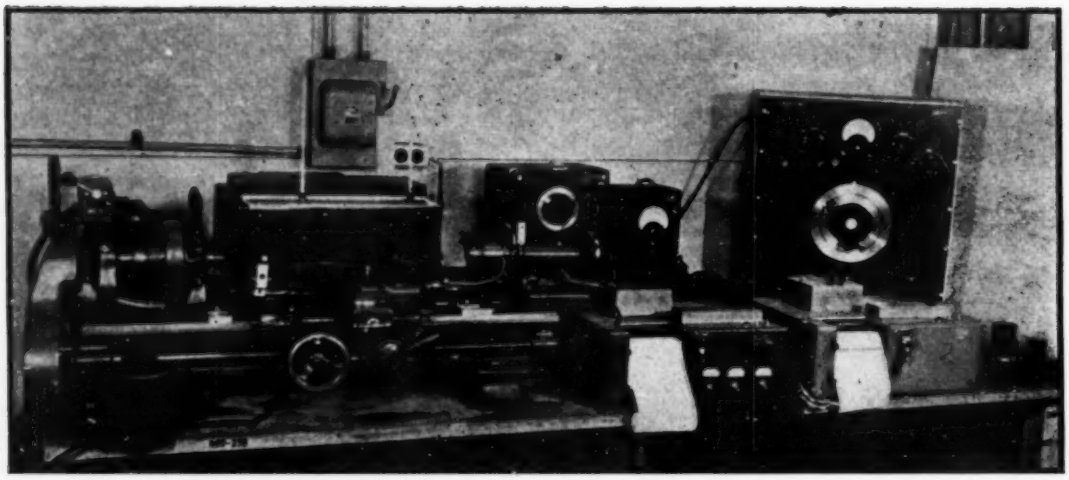


FIG. 5. General view of apparatus for measuring longitudinal-wave transmission in thin rubber strips.

observation of the center frequency can sometimes be made for weaker resonances ("Q" = 1 to 2) at these frequencies. Measurements have been carried out at temperatures from −60°C to 100°C. The equipment is operated in a temperature box. A low power microscope having a graduated ocular scale is fitted into the side of the temperature box to facilitate observation of the vibration of the reed.

The reed method is useful in tracing the temperature dependence of the dynamic modulus at an essentially constant low frequency. Measurements for this purpose are possible when the temperature of the sample is remote from its stiffening temperature. (For every lightly vulcanized rubber, there is a temperature range only a few degrees wide in which the real part of the modulus, measured at low frequencies, increases by a factor of several hundred with reduction of temperature. The loss factor E_2/E_1 has a maximum with respect to temperature in this range.) The reed method also gives useful information regarding the behavior of the modulus in the stiffening-temperature range, but inasmuch as the frequency of vibration, which is proportional to the square root of the modulus, generally changes by a factor of the order of twenty to thirty in the stiffening range, the result cannot be called a constant-frequency determination. Moreover, the large value of the loss factor E_2/E_1 (of the order of unity) in the stiffening range prevents precise measurements or calculations of results in this range. But the loss factor is small for temperatures above and below the stiffening range, at least for vulcanized rubbers, so that the end sections of a modulus-temperature curve through the stiffening range can be determined with precision. Generally, the point-to-point precision of measurements on a given sample is better than the accuracy of the numerical modulus determinations, which is limited by uncertainties in determining the effective dimensions of the sample and its density.

For most materials there has been no indication that the modulus as measured by the reed experiment is a function of the amplitude of vibration. A notable exception to this observation is found in natural rubber, for which the resonance frequency is strongly dependent upon amplitude. The reason for this amplitude dependence requires further elucidation; the effect may occur because in the reed experiment it is impossible to secure differential strains that are small compared to the static strain, since the static strain is zero. The actual deformation is, of course, a combination of extension and compression. It is noteworthy that the actual strains in the reed experiment are very small; for example, the maximum strain is only 0.25 percent under following typical conditions: Reed 1.7 cm in length; thickness 0.08 cm; half-amplitude of vibration at tip, 0.025 cm.

The derivation of the equations for calculation of the complex modulus from the results of the

vibrating-reed experiment is given in Appendix I rather than in the present section, because while the derivation is somewhat lengthy, the details add little to the description of the method. The final equations are

$$E_1/\rho = B_n(l^4/t^2)[\nu_0^2 + (\Delta\nu)^2/2], \quad (21)$$

$$E_2/\rho = B_n(l^4/t^2)\nu_0\Delta\nu. \quad (22)$$

(Error <2 percent for $\Delta\nu/\nu_0(\doteq E_2/E_1)$ <0.2).

Here ρ is the density of the rubber, l the length of the reed, t the thickness of the reed, ν_0 the frequency of resonance in c.p.s., $\Delta\nu$ the band width in c.p.s., and B_n a constant which is associated with the nth mode of vibration. The first mode is that in which there are no nodes along the reed, the second mode is that in which there is one node, and the third mode is that in which there are two nodes. Values of B_n for these first three modes are $B_1=38.4$; $B_2=0.975$; $B_3=0.124$. In practice almost all of the useful experimental data are obtained with the reed vibrating in the first mode, because an appreciable amplitude of vibration in the higher modes is found only in restricted portions of the temperature range in which the damping of the rubber is very small.

IV. LONGITUDINAL-WAVE TRANSMISSION IN THIN STRIPS

The acoustic-wave methods are those in which the quantities measured experimentally are the velocity and attenuation of sound in a sample of the viscoelastic material. The complex modulus may be calculated directly from a knowledge of these quantities and of the density of the sample. In this and the following section two methods for measuring the velocity and attenuation of so-called longitudinal waves in thin strips are described. The strip of viscoelastic material in which the waves are propagated is narrow compared to an acoustic wave-length at all frequencies which are considered. Therefore the wave-propagation constants are determined by the complex Young's modulus and the density. (Further acoustic-wave studies have been devoted to the propagation of bulk waves in rubber at frequencies in the neighborhood of 10^7 c.p.s. Because the bulk wave velocity depends upon both the modulus of compression and the shear modulus, the bulk wave studies are not directly comparable with the experiments described in the present paper, and a detailed description of the bulk wave transmission experiments will be given in a separate paper.)

The method for the study of the transmission of longitudinal waves in thin rubber strips has received a preliminary description in a paper published previously.[7] There it is shown that the real part of Young's modulus and the associated viscosity coefficient are related to the acoustic

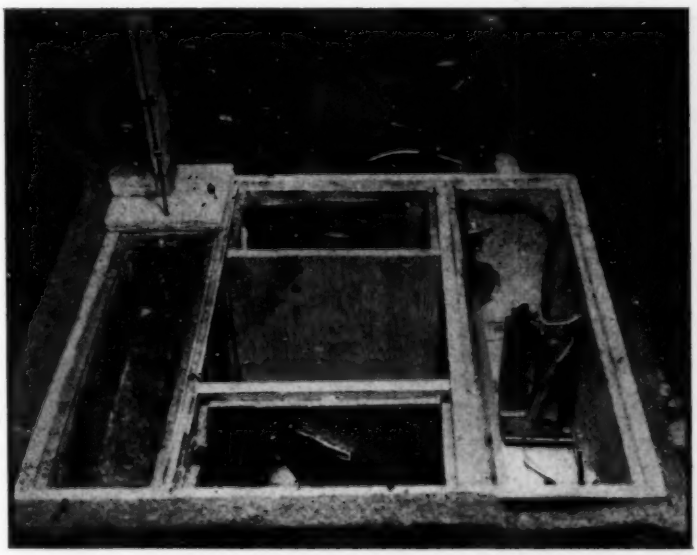


FIG. 6. Top view of temperature box of longitudinal-wave transmission apparatus, cover removed. The rubber strip and the receiving element are in the right-hand compartment.

[7] A. W. Nolle, J. Acous. Soc. Am. 19, 194 (1947).

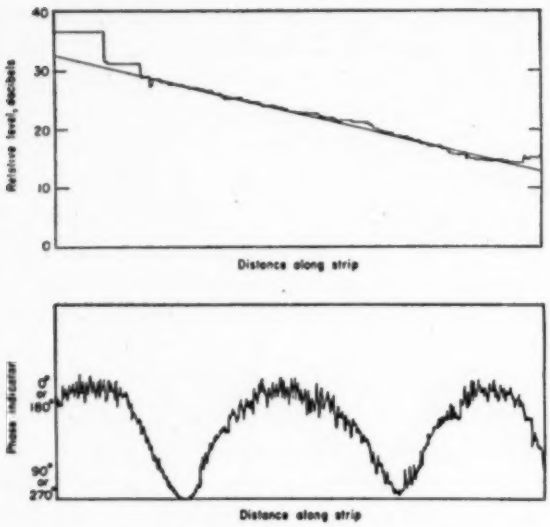


FIG. 7. Above: Typical attenuation record (autographic plot of decibel signal level *vs.* distance) from strip transmission apparatus. The slope of the best straight line drawn through the record, as shown, is proportional to the attenuation in db/cm. Below: Typical phase record from strip transmission apparatus. The distance between successive minima is proportional to a half wave-length.

propagation constants by the equations

$$E_1=\rho c^2\omega^2(\omega^2-\alpha^2c^2)/(\omega^2+\alpha^2c^2)^2, \tag{23}$$

$$\gamma=2\rho\alpha c(\omega c)^2/(\omega^2+\alpha^2c^2)^2, \tag{24}$$

where α, the acoustic attenuation[8] in nepers/cm, and c, the acoustic phase velocity, are the propagation constants measured at the angular frequency ω. The density of the sample is ρ. In practice the attenuation quantity which is measured is the attenuation in db/cm, $L=8.686\alpha$.

For purposes of easy calculation these equations may better be written

$$E_1/\rho c^2=(1-r^2)/(1+r^2)^2, \tag{25}$$

$$E_2/E_1=2r/(1-r^2). \tag{26}$$

The dimensionless parameter r is equal to $L\lambda/54.6$, which is the ratio of the observed attenuation in db per wave-length, $L\lambda$, to 54.6 db, which is the attenuation per wave-length in a purely viscous wave. Thus the maximum value of r is unity. Calculations corresponding to Eqs. (25) and (26) are easily made with the logarithmic nomograph of Fig. 3. A straight edge connecting the observed values of the quantities L and λ in columns I and III intersects the appropriate values of $E_1/\rho c^2$ and of E_2/E_1 in column II. A straight-edge connecting the appropriate values of λ in column III and f (frequency in Kc) in column IV intersects the value of c^2 in column V. A pair of dividers is used to proceed in column V by a distance corresponding to $[\log(E_1/\rho c^2)+\log\rho]$ in the appropriate direction from the point representing c^2, according to the directions on the nomograph. Thus the value of E_1 is obtained.

The experimental method for direct study of acoustic propagation in thin strips consists of recording the logarithmic amplitude and the relative phase, as a function of distance, of the signal received by a pick-up sensitive to longitudinal waves which is in contact with the thin sample strip. During a measurement the pick-up moves uniformly away from a mechanical driving element which is at one end of the strip. Figure 4 is a schematic diagram of the apparatus. The sample is stretched along the center line of a lathe bed between a vibration generator, mounted in the tail stock, and a pulley on a special support near the head stock. The receiver is mounted on the tool post and may be moved uniformly away from the vibration generator by the motor-driven lead screw. The driving element is a phonograph cutting head, which by a suitable linkage produces vibratory motion, in the direction of the strip axis, of a connecting rod which engages a stationer's clip in the end of the sample strip. The receiving element may be either a thin piezoelectric bimorph which contacts the strip lightly, or a phonograph pick-up fitted with a wire stylus which makes contact with the strip. The sample strip is maintained under constant stress by means of a weight and pulley arrangement, or alternatively under constant strain by means of a capstan arrangement.

A general view of the apparatus is shown in Fig. 5. A top view of the temperature box, with the cover removed, appears in Fig. 6. In the compartment at the right may be seen the sample strip and one form of receiving element. The compartment at the left is for "dry ice." Below the "dry ice" compartment, but not visible in the picture, is a compartment which contains an electric heating element. The electric blower (top) continually moves the same air through a

[8] An attenuation of one neper represents a reduction of vibration or wave amplitude by the factor $1/e$.

circuit of the working compartment and the heating-cooling compartments. The division of the air flow between the heating and cooling compartments is controlled by a damper which may be adjusted manually by the rod at the upper left of Fig. 6. In normal operation of the temperature box, the damper is set so that, with the heater turned off, the air stream will assume a temperature slightly lower than the desired working temperature. The heater is then turned on automatically, for such periods as may be required for maintenance of the proper working temperature, by an electronic control circuit whose sensitive element is a "thermistor" (not shown) in the working compartment. A similar temperature box is used in conjunction with each of the other experiments described in this paper.

The amplitude and phase data are plotted continuously on automatic recorder charts moving at constant speed. Typical records are shown in Fig. 7. Ordinarily the acoustic attenuation in the rubber strip is sufficiently large that there is no appreciable reflection of signal from the end of the sample back to the source, and no appreciable reflection from the receiving element back to the source except at the beginning of the attenuation record. For this reason most of the attenuation record approximates a straight-line decrease of logarithmic signal level with distance. The slope of the best straight line drawn through the attenuation record is directly proportional to the attenuation in db/cm. The constant of proportionality is easily found from a knowledge of the relative rates of travel of the receiving element and of the recorder charts.

The phase recorder responds to the output of an electronic circuit whose output goes through a minimum at intervals of 180° phase change. The two alternating voltages which are compared as to phase are the voltage driving the vibration generator and the amplified output of the receiver. Thus the distance between successive minima on the phase record is proportional to one-half of an acoustic wave-length in the sample strip.

The low frequency limit of operation of this apparatus is reached either when the acoustic wave-length in the sample becomes so long that the total phase change over the length of the strip is less than 180°, or when the attenuation in db/cm, which decreases with decreasing frequency, becomes so small that appreciable reflection occurs between the receiver and the driver and the attenuation record is no longer a straight line. In the present apparatus the maximum sample length is about twelve inches, and the low frequency limit for rubber gum vulcanizates is usually about one kilocycle. A high frequency limit for the method is always imposed by attenuation too large for convenient measurement, inasmuch as the attenuation in the sample always increases with increasing frequency. In practice it is difficult to measure attenuation greater than 30–40 db/cm. For some samples the high frequency limit of the measurements is lower than that which would be set by attenuation considerations, and is imposed by the occurrence of undesired wave effects. One such undesired effect is the occurrence of flexural waves. The flexural wave velocity increases with increasing frequency, with the result that while at low frequencies the flexural waves are attenuated rapidly, at sufficiently high frequency flexural waves are observed throughout the length of the strip. (This phenomenon may be understood when it is recognized that the *attenuation per wave-length* of a particular type of wave is a unique function of the parameter E_2/E_1.) The interpretation of the phase and amplitude records which result from a combination of longitudinal and flexural waves is prohibitively difficult. The flexural wave effect is noticeable at lower frequencies in the case of the materials having low dynamic modulus. Another possible difficulty at high frequencies is that the propagation of longitudinal waves is not properly described by the one-dimensional wave equation, which is the basis for Eqs. (23) and (24), in case the longitudinal wave-length is not large com-

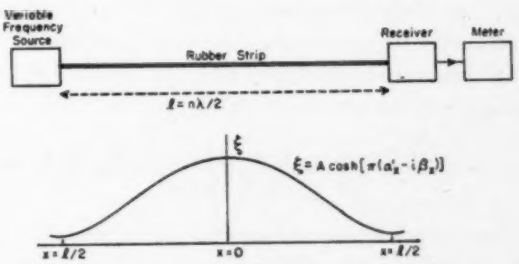


FIG. 8. Above: Schematic diagram of apparatus for study of longitudinal-wave resonance in thin rubber strips. Below: Sketch of amplitude variation with distance, for analysis.

pared to the transverse dimensions of the sample. While the effects of this difficulty can be made to set in at a higher frequency by selecting a thinner and narrower sample, it is found in practice that even with the thinnest practical sample the high frequency limit of measurements may be established by this consideration. Since any of the above factors may set the upper frequency limit of the experiment, depending upon the sound velocity and the attenuation of the sample under test (which are in turn functions of temperature), it is impossible to set a general upper limit. The practical upper limit for various rubber samples may range from 5 to 50 kc, while in the case of other high polymers the method is applicable well above 100 kc. The thickness of the rubber samples used in this apparatus has generally been in the range 0.03 cm to 0.1 cm.

The presence of both flexural and longitudinal waves may result in an attenuation record which does not show a linear decrease of logarithmic signal strength with distance, but which instead shows a periodic perturbation superimposed on the linear record. In the event that attenuation in the sample is sufficiently small to permit the existence of appreciable standing-wave effects, the attenuation record is likewise modulated. Non-linear attenuation records resulting from these two causes may be distinguished from one another on the basis that standing-wave effects produce inflection points in the attenuation record, separated by one-half wave-length, while interference from flexural waves may produce inflection points whose separation bears no simple relation to the longitudinal wave-length. It is possible to make a correction, in the case of standing-wave effects, in order to obtain approximately the correct attenuation constant from the slope of a median line drawn through the record. No simple correction is possible when serious interference from flexural waves exists throughout the attenuation record.

V. LONGITUDINAL-WAVE RESONANCE IN THIN STRIPS

The preceding discussion of sound transmission in thin strips was predicated on the supposition that standing-wave phenomena in the strips are not severe, because of large attenuation in the material. This condition is generally realized when rubber-like materials are tested at frequencies as large as one kilocycle. For somewhat lower frequencies, however, the attenuation per wave-length $L\lambda$ (which is related to the loss parameter E_2/E_1 as shown in the nomograph of Fig. 3) may be sufficiently small that half-wave transmission-line resonance is observed with a sample of suitable length. In this case the complex modulus of the material may be calculated from the band width and the center frequency of the resonance. The experimental technique consists of measuring the amplitude response as a function of frequency at one end of a strip, when the other end is vibrated longitudinally at constant amplitude. The experimental apparatus is indicated schematically in Fig. 8. The transmitting and receiving elements may be phonograph cutting heads connected mechanically to the strip in the same fashion as the driving element of the strip transmission experiment in the preceding section, so as to generate or respond to longitudinal waves. It will be assumed that the receiving element generates a voltage proportional to mechanical displacement. Therefore the receiving element may be either a piezoelectric device of an electrically compensated magnetic device.

This technique has been used with particular success to study the temperature dependence of the modulus of natural rubber at various frequencies from 60 to 300 c.p.s., in the temperature range from 15°C to 70°C. At lower temperatures, in this particular instance, the resonance is not readily observable because of a rapid increase in the loss factor E_2/E_1. The maximum sample length of natural rubber used in the above frequency range was about 25 cm. For most other rubbers it is only at much lower frequencies that the loss factor is small enough to permit readily observable resonance, in which case the method is applicable only to samples of much greater length (especially because the velocity of sound in most other rubbers is greater than in comparable preparations of natural rubber).

A satisfactory analysis of the apparatus for purposes of calculation is easily obtained with the help of certain simplifying assumptions, which are introduced as necessary in the development

which follows. The physical significance of the analysis is indicated in Fig. 8. The important results are given in Eqs. (40) and (41).

The displacement ξ in the transmission line is described by the relation

$$\begin{aligned}\xi_x &= A\cosh[\pi(\alpha_x' - i\beta_x]e^{-i\omega t} \\ &= A[\cosh\pi\alpha_x' \cos\pi\beta_x \\ &\quad - i\sinh\pi\alpha_x' \sin\pi\beta_x]e^{-i\omega t} \quad (27)\end{aligned}$$

where α' and β are the parameters of Morse's[9] hyperbolic tangent impedance method. (Here we follow the usual convention that the value of the physical variable is given by the real part of the complex function.) For the present purpose, α' and β are taken to be

$$\alpha_x' = \alpha'_{l/2} - (\alpha/\pi)[x - (l/2)], \quad (28)$$

$$\beta_x = \beta_{l/2} - [2(x - l/2)/\lambda], \quad (29)$$

where α is the wave attenuation in the material in nepers/cm. Evidently

$$\frac{|\xi_{l/2}|}{|\xi_0|} = \frac{\cosh^2\pi\alpha'_{l/2}\cos^2\pi\beta_{l/2} + \sinh^2\pi\alpha'_{l/2}\sin^2\pi\beta_{l/2}}{\cosh^2\pi\alpha'_{-l/2}\cos^2\pi\beta_{-l/2} + \sinh^2\pi\alpha'_{-l/2}\sin^2\pi\beta_{-l/2}}, \quad (30)$$

where ξ_0 is the displacement at the generator. Suppose that ξ_0 is a constant (does not vary as the transmission frequency characteristic is measured). Also suppose that the mechanical impedance of the receiver which terminates the line is very large compared to the line impedance; thus there is essentially perfect reflection at the termination and $\alpha'_{l/2} \doteq 0$, $\beta_{l/2} \doteq -0.5$, $\beta_{-l/2} \doteq 0.5$. The "Q" of the line must be reasonably high to permit practical measurement of the resonance and to reduce dispersion errors; therefore the total line attenuation $\pi\alpha'$ is not large, and the total variation in $\beta_{-l/2}$ over the band width of the line is small. These assumptions lead to the following simplifications:

$$\begin{aligned}&\cosh\pi\alpha'_{-l/2} \doteq 1; \\ &\cos\pi\beta_{-l/2} = \sin[\pi(0.5 - \beta_{-l/2})] \\ &\qquad \doteq \pi(0.5 - \beta_{-l/2}); \quad (31) \\ &\sin\pi\beta_{-l/2} \doteq 1; \\ &\sinh\pi\alpha'_{-l/2} \doteq \pi\alpha'_{-l/2}.\end{aligned}$$

The numerator of (30) depends upon the impedance of the termination relative to that of the line, but may be considered constant for a given experiment. Thus

$$\xi_{l/2} = \frac{\text{Constant}}{(0.5 - \beta_{-l/2})^2 + \alpha'^2_{-l/2}}. \quad (32)$$

The effect of varying the frequency is (except for a small change in $\alpha'_{-l/2}$) to produce a variation in $\beta_{-l/2}$. Therefore the frequencies at which the amplitude has $1/\sqrt{2}$ times its maximum value correspond to $(0.5 - \beta_{-l/2}) = \pm\alpha'_{-l/2}(\sqrt{2}-1)^{\frac{1}{2}}$. The total change in $\beta_{-l/2}$ corresponding to the total band width is

$$\Delta\beta_{-l/2} = 2\alpha'_{-l/2}(\sqrt{2}-1)^{\frac{1}{2}} = 1.29\alpha'_{-l/2}. \quad (33)$$

But

$$\Delta\beta_{-l/2} = 2l/\Delta\lambda = n\Delta\nu/\nu_0, \quad (34)$$

where n is the mode of vibration in the line (1, 3, 5, etc.), ν_0 is the center frequency for that mode, and $\Delta\nu$ is the total band width in cycles per second. From (33) and (34)

$$1.29\alpha'_{-l/2} = n\Delta\nu/\nu_0. \quad (35)$$

Since $\alpha'_{l/2} \doteq 0$, it follows that

$$\alpha'_{-l/2} \doteq \alpha l/\pi \quad (36)$$

and that

$$\alpha = 0.776(n\pi/l)\Delta\nu/\nu_0. \quad (37)$$

To the accuracy of the previous assumptions, the length of the line is $n\lambda/2$ so that

$$c = 2l\nu_0/n. \quad (38)$$

Equations (37) and (38) give the acoustic propagation constants according to the definitions which were employed in the discussion of the strip transmission method in the previous section. Therefore Eqs. (25) and (26), given in that section, may be used to derive the modulus quantities from the acoustic propagation constants. In terms of the present problem, the parameter r is

$$r = \alpha c/2\pi\nu_0 = 0.776\Delta\nu/\nu_0. \quad (39)$$

It will appear in the next paragraph that the validity of the present method is restricted approximately to the case $r < 0.1$. For this case an

[9] P. M. Morse, *Vibration and Sound* (McGraw-Hill Book Company, Inc., New York, New York, 1936).

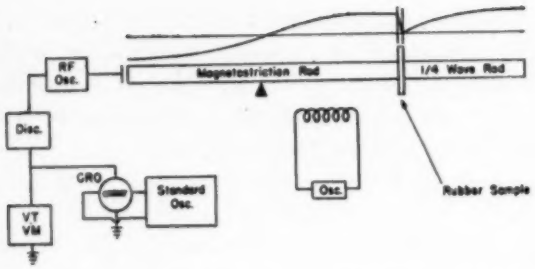


FIG. 9. Schematic diagram of magnetostriction apparatus.

approximate solution of Eqs. (37) and (38) is

$$E_1=\rho c^2=\rho(2l\nu_0/n)^2, \tag{40}$$

$$E_2/E_1=1.55\Delta\nu/\nu_0, \tag{41}$$

where l is the length of the strip, ρ its density, n the mode of vibration (1, 3, 5, etc.), ν the frequency of maximum response in c.p.s., and $\Delta\nu$ the band width in c.p.s., measured between the two frequencies for which the amplitude indicated by the receiver has $1/\sqrt{2}$ times its maximum value.

The approximations in the derivation must be considered further. The original assumptions, that the termination is effectively rigid and that the displacement amplitude of the source is constant, are valid if the equipment is suitably designed. Otherwise the most serious assumption is that $\cosh\pi\alpha'_{-l/2}=1$. The validity of this assumption depends upon the size of the loss factor for the material under investigation, and merits further consideration.

By combining (36), (38), and (39) one finds $\alpha'_{-l/2}\doteq nr$. For a five percent error in the assumption $\cosh\pi\alpha'_{-l/2}=1$, the argument must be restricted to $\pi\alpha'_{-l/2}<0.3$, or $r<0.3/n\pi$. Inserting this condition in Eq. (39) leads to

$$Q=\nu_0/\Delta\nu=0.78/r=0.78n\pi/0.3\doteq 8n. \tag{42}$$

Thus for the first mode the "Q" of the line must not be less than 8, or in other words r must be less than 0.1, corresponding to $E_2/E_1<0.2$, if the error in calculation is to be less than five percent. A more nearly exact solution could be derived in which the error is much less, but the experimental error increases so rapidly with decreasing "Q" that the correction is not worth while. The experimental values of "Q" obtained for the first mode of a transmission line composed of natural rubber gum vulcanizate, at about 100 c.p.s., range from 8 at room temperature to more than 20 at 70°C.

VI. THE MAGNETOSTRICTION METHOD

In the three resonant systems described above, the rubber sample furnishes all of the stiffness in the oscillating system, with the result that the frequency of resonance is proportional to the square root of the real part of the modulus. In the magnetostriction apparatus, however, the resonance which is observed is that of a metal rod in which longitudinal vibrations are excited by magnetostriction, so that the presence of the rubber sample is not a necessary condition for the oscillatory motion to exist. A small change in the frequency of resonance occurs when a rubber sample is forced into contact with the end of the rod. It is this change of resonance frequency which is proportional to the real part of the modulus of the sample. The accompanying change in band width is proportional to the imaginary part of the modulus.

The magnetostriction method is best suited for use in a frequency range from 10 kc to somewhat more than 100 kc. In the lower part of this frequency range it is often more convenient to make modulus measurements by means of the strip transmission technique which is described in Section IV. But although the magnetostriction method is somewhat more tedious, it is capable of providing data under conditions where the strip transmission technique is unsatisfactory, and is generally capable of giving results at higher frequencies than the strip transmission technique.

A schematic diagram of the magnetostriction apparatus is shown in Fig. 9. A small disk sample of rubber is held under compression between a resonant magnetostrictive rod and an antiresonant rod, the latter having half the length of the resonant rod. The antiresonant rod is backed by a layer of spongy rubber, which offers a mechanical impedance at least several times smaller than the characteristic impedance of the rod material. An antiresonant rod has the property that the characteristic impedance of the rod material is the geometric mean between the backing impedance and the impedance which is "seen" at the input end of the rod. Therefore the impedance presented by the antiresonant rod to the sample disk at its left is at least several times

the characteristic impedance of the rod material. For nickel, which is a suitable rod material, the specific characteristic impedance (impedance of an infinitely long rod of cross section 1 cm²) has a value of about 4×10^6 dyne-sec./cm, while the rubber sample is selected as to thickness so that its impedance is about 10^5 to 10^6 at the lowest frequency of resonance of the rod. The rubber sample, therefore, may be regarded as backed by an effectively infinite impedance when, as in this apparatus, the antiresonant rod presents several times its own characteristic impedance. The resonant rod is clamped at its center in a ring of natural rubber, which has little effect upon the system because the impedance of the rod is very large at the center.

The resonant rod is magnetostrictively driven by means of a coil, surrounding the rod, to which is supplied several watts of ac signal. In addition, the coil carries direct current, somewhat greater in magnitude than the peak ac flow, in order to produce magnetic polarization. The relative amplitude of vibration of the rod is measured by means of an electrostatic pick-up plate which is located very near to the free end of the rod. The motion of the rod varies the capacity between this plate and ground, which results in frequency modulation of a high-frequency oscillator (about 10 megacycles). The frequency-modulated oscillator signal is applied to a discriminator, which supplies an output at the signal frequency. This output is amplified in a selective receiver and applied to the vertical input of a cathode-ray oscilloscope.

In the complete apparatus there are two rods, each with a suitable antiresonant rod. The resonant rods are tuned to approximately 12 and 24 kc. The second and third modes of vibration —that is, the third and fifth harmonics—of each rod are also used. The available frequencies are thus 12, 24, 36, 60, 72, and 120 kc. The entire assembly is placed in a temperature control box.

The first essential of the experimental procedure is a careful temperature calibration for each useful resonance of the rods in the absence of a rubber sample. The temperature coefficient of the resonance frequency is large in the case of the pure nickel rods which have been used in most of the work with this technique—about 0.02 percent per centigrade degree.[10] Both the center frequency and the band width are measured in the calibration of the unloaded rods. The change in these two quantities when the sample is in place yields a measure of the viscoelastic coefficients of the sample. Frequency is measured by observing on an oscilloscope the compound pattern which is formed when a mixture of the rod signal and of signal from an accurate low frequency interpolation oscillator is viewed on a standard-frequency sweep. A stationary pattern is obtained by adjustment of the interpolation oscillator frequency. Then the frequency of the "unknown" signal is equal to some multiple of the standard frequency, plus the frequency of the interpolation oscillator. Frequency identification is precise to one or two cycles per second, and is not a limiting factor in the over-all precision of the experiment. Amplitude is observed directly on the oscilloscope screen.

The real and imaginary parts of Young's modulus are found from the following relations:

$$E_1=t(4\pi^2\rho l)\nu_c(\nu_c-\nu_0), \tag{43}$$

$$E_2=t(4\pi^2\rho l)\nu_c(\Delta\nu-\Delta\nu_0)/2. \tag{44}$$

(Error <5 percent if $\Delta\nu/\nu_c<0.1$ and $(\nu_c-\nu_0)/\nu_0<0.05$.)

Here t is the thickness of the sample, ν_c is the center frequency of the loaded resonance, ν_0 is the frequency of the unloaded resonance, $\Delta\nu$ is the band width[5] of the loaded rod, $\Delta\nu_0$ is the band width of the unloaded rod. Frequencies and band widths are in cycles/sec. The quantity $4\pi^2\rho l$ is a constant for a particular rod; ρ is the density of the rod material and l is the length of the rod. The density and sound velocity of nickel, in particular, vary from one sample to another. Consequently, it is desirable that the density of the particular specimen of nickel be measured rather

[10] More recent work indicates that satisfactory results may be obtained with rods composed of a stainless steel composition intended for use in low temperature-coefficient magnetically driven tuning forks. The temperature coefficient of the resonance frequency is only 0.002 percent per C° for these rods. In addition, the stainless steel rods have a "Q" between 5000 and 10,000, about 10 times the value found for unloaded nickel rods. These features greatly facilitate the frequency-temperature calibration of the unloaded rods. On the other hand, the stainless steel alloy has a magnetostriction constant smaller than that for nickel, which in practice makes it more difficult to secure measurements at the higher harmonics, when the rod is loaded with a rubber sample, than with nickel rods.

TABLE I.

Method	Nominal frequency range	Static deformation	Special advantages	Limitations
Rocking-beam oscillator.	0.1 to 25 c.p.s.	Elongation, 5 percent to upper limit which depends on sample stiffness and beam weight.	Precise indication of temperature behavior of modulus at nearly constant frequency. Lower frequency of operation than other methods.	Not suitable below temperatures for which $E_2/E_1=0.2$. Fairly tedious.
Vibrating reed.	10 to 500 c.p.s.	None.	Study of low temperature stiffening at low frequencies.	Frequency adjusted only by trimming reed. Data inaccurate for $E_2/E_1>0.2$ and observations impossible for $E_2/E_1>1$. Tedious near stiffening temperature.
Strip transmission.	1 to 40 kc (depends on material and temperature).	Elongation, 2 percent to strength limit of clamping arrangement.	Automatic operation. Continuous frequency coverage at fixed temperature. Convenient.	Not suitable for cold stiff material nor for attenuation >40 db/cm nor for $\lambda<0.25$ cm. Possible flexural wave interference in low-modulus materials.
Strip resonance.	Low limit set by sample $=\lambda/2$. Upper limit set by attenuation (300 c.p.s. with natural rubber).	Elongation, 1 or 2 percent to strength limit of clamping arrangement.	Precise indication of modulus *vs.* temperature at nearly constant frequency. Convenient.	Not suitable below certain temperature or above certain frequency. Limit set by requirement $E_2/E_1<0.2$.
Magnetostriction.	12 to 120 Kc	Compression, up to 10 percent or more, depending upon stiffness of sample and strength of apparatus.	No E_2/E_1 restriction. Wide temperature range. Extends frequency range of strip method. Very small sample.	Sample size and thickness must be controlled so $Q>10$, frequency deviation <5 percent; sample also thin compared to λ. Frequency not continuously adjustable. Tedious.

than that a handbook value be used. It is assumed that the rubber sample covers the entire end surface of the rod. The derivation of these relations is given in Appendix II.

The magnetostriction method is advantageous because in principle it is always possible by this method to obtain viscoelastic measurements at the particular frequencies determined by the rods, no matter how large the loss factor E_2/E_1. In practice certain precautions must be observed, however. The sample must be sufficiently thin that the loading on the rod is great enough to be experimentally detectable with reasonable precision. On the other hand, the sample must not load the rod too severely; otherwise the amplitude of vibration may be reduced below the limit of measurement, or the change in frequency of resonance may be too large in relation to certain approximations which are made in the formulas for calculation. (See Appendix II.) The dynamic modulus of actual rubber-like materials varies by a factor of 100 to 1000 with changes in temperature. As a result, it is necessary to provide the samples in two or three thicknesses in order to study modulus over all of the important range of temperature variation. For example, in the study of a soft gum vulcanizate of Buna-N rubber (Hycar OR-15), it was necessary to use a sample 0.18 cm thick for temperatures below 10°C, a sample 0.09 cm thick in the neighborhood of 10°C, and a sample 0.025 cm thick for temperatures from 10° to 50°C.

Another precaution must be observed in regard to the wave-length of sound in the sample. It is assumed in the calculations for this experiment that the sample thickness is small compared to a wave-length—say, not greater than $\frac{1}{12}$ wave-length. (Otherwise the sample cannot be considered as a simple damped spring.) The experimental modulus data must always be checked to insure that this criterion has not been violated. Fortunately, the requirements of maintaining proper loading of the rod and of maintaining a sample thickness small compared to a wave-length are not at variance with one another for rubber-like materials which are loaded moder-

ately or not at all with filler. For much stiffer materials, the loading on the rod due to a $\frac{1}{12}$ wave-length or thinner sample would be excessive, although the loading could be reduced by decreasing the area of the sample. A further wave-length consideration involves the requirement that the wave-length of sound in the sample be several times greater than the diameter of the rod. If this requirement is violated seriously, the sample will not expand laterally to the extent which is necessary for a true measurement of Young's modulus. This difficulty may be averted by perforating the sample with an array of small holes. In the existing apparatus, the rod diameter is 0.25 inch and the sample is perforated with 0.03-inch diameter holes on a 0.05-inch square lattice which fills an area 0.25 inch square. There has been no indication that the measured modulus is not a Young's modulus. When perforated samples are used, the true Young's modulus is equal to the measured Young's modulus multiplied by $1/(1-f)$, where f is the fraction of the surface area which is occupied by perforation.

The precision of the magnetostriction method is somewhat limited; the uncertainty of the measurements is frequently as large as 10 to 20 percent. One important source of uncertainty is the difficulty of measuring accurately the center frequency and the band width for the broad resonance which occurs with the rubber sample in place. Another difficulty arises from uncertainty in the calibration of the unloaded rods. Unfortunately, the shift in frequency arising from a sample of given stiffness is smaller for the higher harmonics of a particular rod, even though the difficulty in securing accurate calibration is greater at the higher frequencies. Some uncertainty in the final data also results from imperfect temperature regulation. The experimental measurements are inherently time-consuming, and the rate of securing data becomes so slow, if optimum thermal equilibrium is awaited between readings, that a practical compromise between reduction of thermal errors and rapid yielding of data is usually desirable.

The dynamic modulus as determined by the magnetostriction method is closely but not exactly comparable with that obtained by the longitudinal-wave methods. For many rubber compounds, the dynamic modulus is appreciably affected by the static deformation of the material. In the longitudinal-wave methods the static deformation is an elongation of five percent or more, whereas in the magnetostriction method the static deformation is a compression, produced by a force which is adjustable in steps up to about 6 kg/cm².

VII. SUMMARY OF EXPERIMENTAL METHODS

In order to afford a convenient summary of the experimental methods described above, these methods are listed in Table I along with their outstanding advantages and limitations.

The previous sections have dealt with the general principles of the experimental methods rather than with the application of the methods to the study of particular rubber compounds. Consequently little specific information has been given concerning the range of temperatures in which the methods are applicable. While a temperature box of the type described in Section IV is capable of operating between the approximate limits −60°C and 100°C, the characteristics of a particular rubber may be such that certain of the experimental methods are useful through only a portion of this temperature range. Moreover, most rubbers undergo irreversible changes at temperatures exceeding some 80°C, so that measurements at temperatures higher than this are not ordinarily desirable. Some idea of the actual temperature ranges of the various methods in a practical case may be obtained from the following paragraphs, in which the applicability of the various methods is discussed with reference to a specific Buna-N rubber compound (Hycar OR-15. 100; ZnO, 5; S, 2; mercaptobenzothiazole, 2, all in parts by weight; 30 min. cure, 149°C). The stiffening temperature of this compound at 100 c.p.s. is about 0°C.

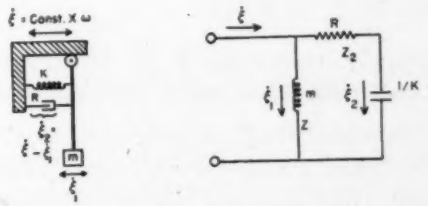


FIG. 10. Lumped-constant mechanical representation of vibrating reed, and equivalent electrical network.

Rocking-beam oscillator measurements at about 0.1 c.p.s. may be obtained at temperatures above 0°C. Because of the increase of the loss factor E_2/E_1 with increasing frequency, measurements are not possible below 30°C if the frequency is increased to 25 c.p.s.

Vibrating-reed measurements with this compound may be performed at any attainable temperature, but the data in the range −5°C to 5°C are not precise because the loss factor E_2/E_1 is large in this range. In a typical set of measurements the frequency may be about 25 c.p.s. at temperatures above the stiffening range and about 500 c.p.s. at temperatures below the stiffening range.

The strip transmission apparatus yields data for this compound at temperatures of 5°C and above. The frequency range is approximately 1 kc to 20 kc. The existing apparatus is not used at temperatures above about 40°C, in order to prevent damage to the Rochelle salt crystal receiving element, but there is no characteristic of the rubber which would in itself prevent measurements at higher temperatures.

The strip resonance method is useful in the case of this compound only at frequencies below 100 c.p.s., and at temperatures above 20° to 30°C. This method is inconvenient for a Buna-N compound because a sample of the order of one meter long is required. The strip resonance method is better suited to rubbers having low dynamic modulus and low loss, as for example natural rubber or GR-S.

The magnetostriction method will supply data at any temperature attainable with the apparatus, provided that samples are supplied in suitable thickness for the different ranges of temperature (see Section VI).

In conclusion the author is pleased to acknowledge the kind advice and guidance of Professor R. H. Bolt, and the invaluable laboratory assistance of Mr. S. C. Mowry.

APPENDIX I. DERIVATION OF FORMULAS FOR THE VIBRATING-REED EXPERIMENT

The mathematical analysis of flexural vibrations is rather involved, even in cases where the boundary conditions are simple and there is no loss in the material which transmits the vibrations. The case of the viscoelastic reed is complicated both by loss in the material and by forced motion at one end of the vibrating member. Consequently the problem will be solved by an equivalent-circuit artifice rather than by rigorous analysis.

The frequencies of free vibration of a bar with one clamped end, one free end, are given by Morse[9] (in slightly different form) as

$$\omega_n = A_n(E)^{\frac{1}{2}}, \tag{45}$$

$$A_n = \pi^2 t \beta_n^2/(12\rho)^{\frac{1}{2}} l^2, \tag{46}$$

where E is the Young's modulus of the bar material, l is the length in cm of the bar, t is the thickness of the bar, and ρ is the density of the bar material. The factor $t/(12)^{\frac{1}{2}}$ is the radius of gyration of the cross section of the bar about an axis which is perpendicular to the length of the bar and to the direction of vibration; the factor as given applies to a bar of rectangular cross section. β_n is a constant associated with the nth mode of vibration. Values for the first three modes are: $\beta_1=0.597$; $\beta_2=1.494$; $\beta_3=2.500$.

A well-known result of the general theory of vibrating systems is that the amplitude-frequency response of a distributed system, near one of its resonances, is to a first approximation the same as that of a simple resonant system having "lumped" constants but operating near its frequency of resonance. Thus the forced-vibration behavior of the viscoelastic reed near resonance can be analyzed by considering an equivalent circuit having lumped constants which are identified with the constants of the reed.

An equivalent circuit for the viscoelastic reed, driven by motion of the clamp, is shown in both mechanical and electrical form in Fig. 10. The displacement amplitude of the driver ξ is independent of frequency (true below driver resonance), or, in other words, the velocity amplitude $\dot{\xi}$ is proportional to frequency. The velocity amplitude in the mechanical circuit is regarded as current in the electrical circuit. The constants of the equivalent circuit may be identified with the constants of the reed as follows:

m corresponds to $1/A_n^2$;
k corresponds to E_1;
R corresponds to γ, or E_2/ω.

The complex modulus $E^*=E_1-iE_2$ has been substituted here for the simple modulus E.

The displacement at the free end of the reed is $\dot{\xi}_1/\omega$. By applying the usual procedures of ac circuit analysis, one may verify the following expression for the square of the magnitude of this quantity:

$$|\dot{\xi}_1/\omega|^2=\text{Constant}\times\frac{\omega_1^2\omega^2+\omega_1^4Q^2}{\omega_1^2\omega^2+Q^2(\omega^2-\omega_1^2)^2}, \quad (47)$$

where ω_1 denotes $(k/m)^{\frac{1}{2}}$ and Q denotes $m\omega_1/R$. The frequency of maximum response is obtained by differentiating the above expression with respect to ω and setting the result equal to zero. The square of the frequency of maximum response, ω_0^2, is given by a quadratic equation whose solution reduces by binomial theorem approximations for $Q\gg1$ to

$$\omega_0^2=\omega_1^2(1-1/2Q^2). \quad (48)$$

In addition to the frequency of maximum response, the band width is required. Let $\omega_0(1\pm\Delta\omega/2\omega_0)$ be the frequencies at which $|\dot{\xi}_1/\omega|^2$ has half the value found at frequency ω_0. This condition, applied to Eq. (47), gives rise to the equation

$$2\frac{(1+1/2Q^2)(1\pm\Delta\omega/\omega_0)+Q^2(1+1/2Q^2)^2}{(1+1/2Q)^2(1\pm\Delta\omega/\omega_0)+Q^2(\Delta\omega/\omega_0\pm1/2Q^2)^2}=\frac{(1+1/2Q^2)+Q^2(1+1/2Q^2)^2}{(1+1/2Q^2)+1/4Q^2} \quad (49)$$

which has been simplified by the relation $\omega_1^2\doteq\omega_0^2(1+1/2Q^2)$. An approximate solution for this equation is

$$1/Q=(\Delta\omega/\omega_0)[1-(3/8)(\Delta\omega/\omega_0)^2]. \quad (50)$$

To the degree of the approximations, the same solution is given by either the plus or the minus signs in (49).

The above solutions were obtained under the supposition that the resistance R in the resonant system is independent of frequency. In actual rubber-like materials, neither the viscous (resistive) coefficient γ nor the product $\gamma\omega=E_2$ is independent of frequency. But a different value of the coefficient of $(\Delta\omega/\omega_0)^2$ in (50) is obtained if the resistance varies with frequency. A practical solution to this dilemna is to drop the second-order term in (50) altogether and to assign a corresponding error to the calculation. The absence of the second-order term introduces an error of the order of $1/2Q^2$, or 2 percent for $Q=5$. To a first-order approximation Q is equal to $\omega_0/\Delta\omega$.

By replacing $Q=m\omega_1/R$ and $\omega_1=(k/m)^{\frac{1}{2}}$ with the reed constants which are identified with these quantities, inserting the expression for A_n, and replacing the radian frequency ω by $2\pi\nu$, one obtains from (48) and (50) the expressions (21) and (22) for E_1 and E_2.

APPENDIX II. DERIVATION OF FORMULAS FOR THE MAGNETOSTRICTION EXPERIMENT

The forced-vibration resonance of a rod which is loaded at one end is most easily studied through a mechanical impedance analysis. Let the rod have a cross section of A cm^2, and a length of l cm. The area of the rubber sample is also A cm^2. A sinusoidal force having a constant amplitude of F dynes acts at the end of the rod, where the rubber is in contact with the rod. The impedance Z_m associated with the force is the impedance of the rod plus the impedance of the rubber sample. Accordingly the velocity at the loaded end of the rod is

$$\dot{\xi}=F/Z_m=F/[-iA\rho c\tan(2\pi l/\lambda)+i(E_1-iE_2)(A/\omega t)] \quad (51)$$

where

$$\tan(2\pi l/\lambda)=\tan[n\pi(\omega-\omega0_n)/\omega0_n]\doteq n\pi(\omega-\omega0_n)/\omega0_n. \quad (52)$$

Here t is the thickness of the rubber sample, ω is the frequency of the impressed vibration and $\omega0_n$ is the frequency of resonance of the unloaded rod in the nth mode of vibration ($n=1, 3, 5$, etc.) The actual frequency is presumed to differ by no more than a few percent from one of the frequencies of free vibration. With this simplification (51) becomes, in terms of the amplitude $|\xi|=|\dot{\xi}|/\omega$,

$$|\xi|=\left|\frac{tF/A}{E_2-i\omega(\rho cn\pi[(\omega-\omega0_n)/\omega0_n]t-E_1/\omega)}\right|. \quad (53)$$

If E_2 is presumed to be sensibly independent of frequency over a small range, the frequency of maximum amplitude is that for which the second term in the denominator is zero. Thus the reso-

nance condition gives

$$E_1 = t\rho cn\pi(\omega c_n)(\omega c_n - \omega 0_n)/\omega 0_n, \qquad (54)$$

where ωc_n is the frequency of maximum amplitude response in the nth mode for the rod loaded with the rubber sample.

The two frequencies whose difference is the band width are those for which the real and imaginary terms in the denominator of (53) are equal; their difference is $\Delta\omega$. Thus it is possible to derive from (53) the following equation for $\Delta\omega$:

$$\Delta\omega_n = (\omega 0_n/2)\{[1+4(E_1+E_2)/t\rho cn\pi\omega 0_n]^{\frac{1}{2}} - [1+4(E_1-E_2)/t\rho cn\pi\omega 0_n]^{\frac{1}{2}}\}. \qquad (55)$$

The use of binomial-theorem approximations for the condition that the second terms in the radicals are small compared to unity, and the use of the relation (54) for E_1, lead to

$$E_2 = t\rho cn\pi(1+2D)\Delta\omega_n/2 \qquad (56)$$

as a solution for (55), where the frequency deviation D is defined as

$$D = (\omega c_n - \omega 0_n)/\omega 0_n. \qquad (57)$$

Introduction of the relation

$$c = 2\nu_0 l/n \qquad (58)$$

permits (54) and (56) to be transformed to the following forms:

$$E_1 = tK\nu_c(\nu_c - \nu_0), \qquad (59)$$

$$E_2 = tK\nu_c(1+D)\Delta\nu/2 \qquad (60)$$

in which the subscripts n have been dropped and the radian frequency ω has been replaced by $2\pi\nu$.

Since approximations have been introduced in order to permit solutions simple enough for practical use, it is necessary to gain some notion of the error involved in these approximations. This error is a function of the frequency deviation D. For $D=0.05$, the tangent approximation (52) makes the final results for both E_1 and E_2 too small by 1 percent, while the approximations made in deriving (56) from (55) render the E_2 result too small by an additional 1 percent.

In practice, better experimental precision is obtained by taking the center frequency of resonance to be the average of the two half-power frequencies. The error introduced by this practice is a function of $(\Delta\nu/\nu_c)$. It may be shown that a first-order correction for this effect consists of multiplying the value of E_1, as found from Eq. (59), by $[1+(\Delta\nu)^2/16D\nu_c^2]$. This correction is small and need not be used in practice. The effect of omitting this correction is to make the result for E_1 too small by about 1 percent for the case $D=0.05$, $\Delta\nu/\nu_c=0.1$.

The most serious error in calculation results from a feature of the equipment which is not considered in the above analysis—namely, that the vibration amplitude is measured at the end of the rod opposite to that at which the driving force is applied. To correct for this condition one should multiply the amplitude equation (53) by $1/\cos(\omega l/c)$. The algebraic examination of this correction is too involved to justify the space required for its presentation. For the case $D=0.05$, $\Delta\nu/\nu_c=0.1$, this error makes the result for E_1 about 5 percent too large, and the result for E_2 about 10 percent too large. In order to correct partially for this relatively large error in E_2 it is expedient to drop the term D in Eq. (60), with the result that the final equations for calculation assume the form of (43), (44). When calculations are made from the latter equations the net error of calculation for either E_1 or E_2 is 3 percent for the case $D=0.05$, $\Delta\nu/\nu_c=0.1$. The error in E_2 reaches 7 percent in the case $\Delta\nu/\nu_c \ll 0.1$, $D=0.05$, but this condition is not encountered with actual rubber-like materials except at temperatures well below the stiffening range. The errors decrease rapidly as both D and $\Delta\nu/\nu$ decrease.

Any loss inherent in the unloaded resonant rod is effectively added to the loss introduced by the sample. Therefore, if the band width of the unloaded rod is not negligible in comparison to the band width observed with the sample in place, the quantity $\Delta\nu$ to be used in Eq. (60) should be obtained from the relation

$$\Delta\nu = \Delta\nu_{\text{loaded}} - \Delta\nu_{\text{unloaded}} \qquad (61)$$

as indicated by the form of (44).

Nucleation

J. C. Fisher, J. H. Hollomon, and D. Turnbull
Research Associates, General Electric Research Laboratory, Schenectady, New York
(Received February 4, 1948)

Based on Turnbull's concept of the importance of transient nucleation during phase transformations in solids, a general theory of nucleation has been developed. The nucleation mechanism for reactions where stable nuclei appear only during cooling or plastic deformation has been clarified. The general theory makes possible quantitative description of the rates of nucleation of all phase transformations as well as of fracture cracks, slip bands, and magnetic transformations in solids.

MANY phase transformations have been discussed satisfactorily in terms of the concepts of nucleation and growth. Particles of a new phase are assumed to form, and to change their size in the parent phase, by statistical fluctuations. The particles that reach a critical size required for continuous growth at the expense of the parent matrix grow and attain appreciable dimensions. The rate of appearance of continuously growing regions is termed the rate of nucleation, and this rate together with the rate of growth of stable regions of the new phase completely determines the kinetics of a given transformation.

The concepts of nucleation and of growth have adequately described the isothermal phase transformations to which they have been applied.[1,2] However, these concepts have not yet successfully been used to describe the kinetics of a number of transformations that are not observed to take place isothermally. A typical reaction of this type is the austenite to martensite transformation in steels, for which the new phase appears only during cooling of the parent phase. If the cooling is interrupted, transformation ceases; yet had the cooling continued, transformation would also have continued.

The apparent inapplicability of nucleation theory to martensite and martensite-like reactions has led to confusion and error in interpreting transformations of this kind. It is important to develop a consistent and quantitative theory of the kinetics of all types of phase transformations. The foundation for such a theory can be found in Turnbull's recent contribution to the theory of nucleation.[3]

Particles of a new phase that exceed the critical size required for continuous growth are commonly called nuclei. Particles of subcritical size will be called embryos in order to differentiate them from nuclei. Turnbull has discussed the way in which the sizes of embryos change by statistical fluctuations. Nuclei do not leap into existence with a single fluctuation; rather, they arise from embryos that change their sizes continuously at finite rates by losing or gaining atoms one at a time from the surrounding matrix. An important consequence of this idea is that the transient concentration of embryos of every size can differ greatly from the equilibrium or steady state concentration whenever insufficient time has been allowed for steady-state conditions to be realized.

Although transformations such as condensation from the vapor have been discussed successfully in terms of the steady-state rate of nucleation,[4,5] Turnbull has emphasized the importance of the transient rate for transformations in solids, where most reactions are completed before the steady state has been attained. For example, the incubation period frequently observed in solid-solid phase transformations represents the finite time required for embryos to reach critical size.

It is the purpose of this paper to demonstrate that as a result of Turnbull's contribution the concepts of nucleation and growth lead to a

[1] W. A. Johnson and R. F. Mehl, Trans. A.I.M.M.E. **135**, 416 (1939).
[2] R. F. Mehl and L. K. Jetter, "Age hardening of metals," Am. Soc. Metals Symposium, 342–417 (1939).
[3] D. Turnbull, "Transient nucleation" (to be published).
[4] M. Volmer and H. Flood, Zeits. f. Physik. Chemie **170A**, 273 (1934).
[5] R. Becker and W. Döring, Ann. d. Physik [5] **24**, 719 (1935).

comprehensive and quantitative mechanism for the kinetics of *all* phase transformations.

BASIC THEORY

If i contiguous atoms (or molecules) of a parent phase α are transformed to an embryo (or nucleus) of phase β, the free energy associated with the material in the neighborhood of the embryo is altered. A new $\alpha-\beta$-interface is created, and interfacial free energy appears. The bulk-free energy of phase β is generally different from that of phase α, leading to a second free energy change. Further, the specific volume of phase β will generally differ from that of phase α, producing strain energy.

The free energy change in the neighborhood of an embryo containing i atoms can be written as

$$\Delta F_i = Ai^{\frac{1}{3}} + Bi + Ci, \tag{1}$$

where A is proportional to the interfacial free energy, B is proportional to the bulk free energy difference between phases β and α in the absence of stress and surfaces, and C is proportional to strain energy. There are no disposable parameters in Eq. (1). A is the product of the $\alpha-\beta$-interfacial energy in ergs per square centimeter and a conversion factor, depending upon the volume per atom and the shape of the embryo, for changing $i^{\frac{1}{3}}$ to square centimeters. B is the free energy change in ergs corresponding to the transformation of one mole of phase α to phase β in the absence of stress and surfaces, multiplied by $1/N$ where N is the number of molecules per mole. C is the product of $1/N$ and the strain energy, computed by elastic theory, accompanying the transformation of one mole of phase α to phase β in a matrix of α, where the β-phase has a shape geometrically similar to that of the embryo. The embryo shape, upon which both A and C depend, is determined by minimizing ΔF_i with respect to shape.

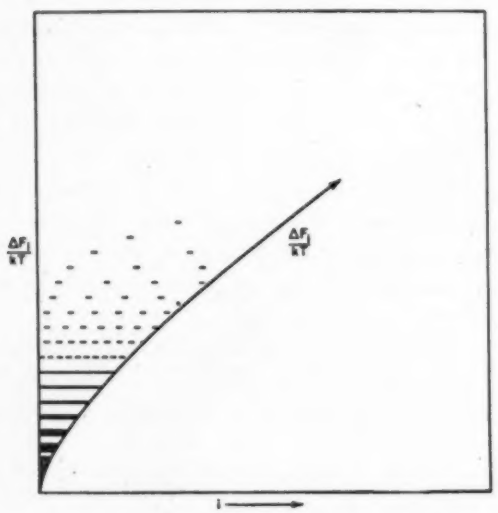


FIG. 1. $\Delta F_i/kT$ *versus* i for transformation to an unstable phase.

The equilibrium number of embryos containing i atoms can be obtained by minimizing the free energy of the system with respect to the number n_i of embryos. The free energy change of the system upon introducing n_i embryos is

$$\Delta F = n_i \Delta F_i - T\Delta S_{n_i} \tag{2}$$

where ΔS_{n_i} is the entropy of mixing n_i embryos with the remaining particles. Assuming the number of embryos of all sizes to be very much smaller than the number of atoms of untransformed α, the expression for the entropy of mixing simplifies to

$$\Delta S_{n_i} = kn_i[1 - \log(n_i/n)]. \tag{3}$$

Setting $\partial \Delta F/\partial n_i$ equal to zero,

$$\partial \Delta F/\partial n_i = \Delta F_i + kT \log(n_i/n) = 0 \tag{4}$$

and the equilibrium number of embryos is

$$n_i/n = e^{-\Delta F_i/kT}. \tag{5}$$

Equation (5) is valid only for ideal mixing when the number of atoms in the embryos is very small, for otherwise the entropy of mixing is not given by Eq. (3).

It is necessary to consider in more detail the mechanism by which embryos of a given size appear. They are produced by a number of bimolecular reactions, where each embryo is considered to be a molecule. The chain of reactions is

$$\left.\begin{aligned} m\alpha_1 &\rightleftharpoons \beta_m, \\ \beta_m + \alpha_1 &\rightleftharpoons \beta_{m+1}, \\ \beta_{m+1} + \alpha_1 &\rightleftharpoons \beta_{m+2}, \\ \beta_{m+2} + \alpha_1 &\rightleftharpoons \beta_{m+3}, \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned}\right\} \tag{6}$$

where α_1 is a single atom of phase α, β_i is an embryo of phase β containing i atoms, and m is

the number of atoms contained in the smallest embryo of β.

In general, the number of embryos of each size β_i will approach an equilibrium or steady-state concentration if the external conditions such as temperature and pressure are constant. The steady-state concentration of embryos is the limiting number per unit of untransformed α. If the transformation is not proceeding, the steady state is an equilibrium state. The rapidity with which the steady-state condition is attained will decrease as the temperature is lowered, since all the bimolecular reactions[6] are thereby retarded.

Frequently it is desirable to calculate the steady-state concentration of β-embryos in phase α during the transformation $\alpha \rightarrow \beta$. The steady state rate of nucleation I_β for phase β is the number of growing regions of β which appear per unit volume of untransformed α per unit time. I_β is also the excess of embryos β_i which add one α atom to become β_{i+1}, over β_{i+1} which lose one atom to become β_i. If $I_\beta \ll n_i$, the steady state concentration of embryos β_i will not differ appreciably from

$$n_i/n = e^{-\Delta F_i/kT}, \tag{5a}$$

as given by Eq. (5), where n represents the number of untransformed atoms, and n_i the corresponding number of embryos β_i. The conditions under which the above equation applies are clarified below.

THERMAL NUCLEATION

The free energy change associated with the transformation of a region containing i atoms is written above as

$$\Delta F_i = Ai^{\frac{1}{2}} + Bi + Ci,$$

where A, B, and C are functions of temperature. A and C, representing an interfacial energy and a strain energy, are positive. B is positive when the parent phase is the more stable, and is negative when the parent phase is the less stable.

Consider first the transformation $\alpha \rightarrow \beta$ at a temperature for which α is the more stable phase. The bulk free energy term B is positive, and all three terms of ΔF_i increase as i increases. The curve of $\Delta F_i/kT$ *versus* i also increases without limit as i increases, as shown in Fig. 1. The weights of the horizontal lines in Fig. 1 vary as $e^{-\Delta F_i/kT}$ at the corresponding values of $\Delta F_i/kT$; the weight of each horizontal line therefore represents the relative equilibrium number n_i of embryos of phase β corresponding to the value of i at which that line intersects the curve of $\Delta F_i/kT$ *versus* i. The interpretation of Fig. 1 is further illustrated by Fig. 2, where the curve of n_i *versus* i is drawn so that the value of $n_i \sim e^{-\Delta F_i/kT}$ is always proportional to the weight of the corresponding horizontal line in the plot of $\Delta F_i/kT$ *versus* i.

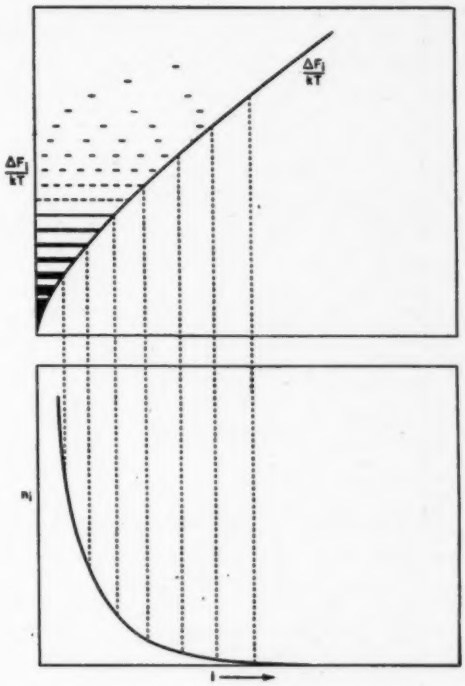


FIG. 2. Relative numbers of embryos for transformation to an unstable phase.

Figures 1 and 2 show clearly that the equilibrium number of embryos of a relatively unstable phase β in a more stable phase α decreases rapidly with increasing embryo size. No nuclei of phase β ever form, and the "transformation" to phase β ceases with the formation of a few embryos of the new phase. A given embryo exists only temporarily, shrinking to α after a brief existence as β.[6]

Consider next the transformation $\alpha \rightarrow \beta$ at a temperature for which β is the more stable phase. The second term of the expression $Ai^{\frac{1}{2}} + Bi + Ci$

[6] J. Frenkel, J. Chem. Phys. **7**, 200, 538 (1939).

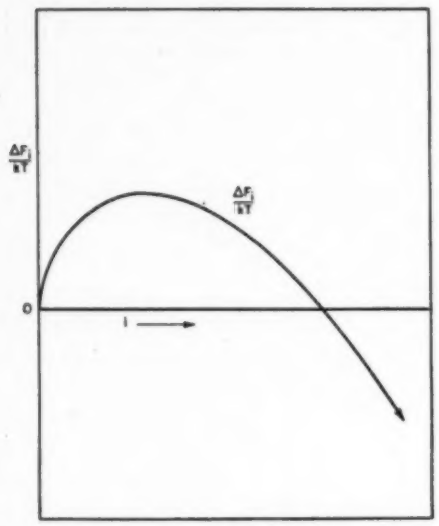


FIG. 3. $\Delta F_i/kT$ *versus* i for transformation to a more stable phase.

is negative, and $\Delta F_i/kT$ has a maximum at a large value of i. Large embryos will have a shape sufficiently plate-like that $|C|<|B|$ and the quantity $(B+C)$ is negative, assuring that the curve of $\Delta F_i/kT$ *versus* i will ultimately pass through a maximum even though the plate-like shape has greatly increased the interfacial energy term A. Figure 3 shows that the corresponding curve of $\Delta F_i/kT$ *versus* i passes through a maximum and then decreases without limit. Imagine that a vertical free energy barrier exists at the maximum of the $\Delta F_i/kT$ *versus* i curve. With this restriction an equilibrium distribution of embryos exists as shown in Fig. 4. The number of embryos of each size is given by the equilibrium relationship $n_i/n=e^{-\Delta F_i/kT}$.

If now the vertical free energy barrier of Fig. 4 is removed, some embryos will spill over the free energy hill, become nuclei, and grow. A steady state will be approached, as shown in Fig. 5, when the rate of nucleation (i.e. the net rate of the process embryo→nucleus, or the net rate of appearance of growing regions of β per unit time per unit volume of untransformed α) becomes constant.

It is evident that the steady state concentration of embryos does not differ appreciably from the equilibrium number

$$n_i/n=e^{-\Delta F_i/kT},$$

calculated assuming a vertical barrier at the critical nucleus size. Only for embryos very near the critical size, or obviously for nuclei exceeding the critical size, is the above expression appreciably in error. Further, the rate of nucleation, or current of embryos passing over the free energy hill, is proportional to the equilibrium number of critical size embryos

$$n_i^*/n=e^{-\Delta F_i^*/kT},$$

calculated assuming the vertical barrier. It should

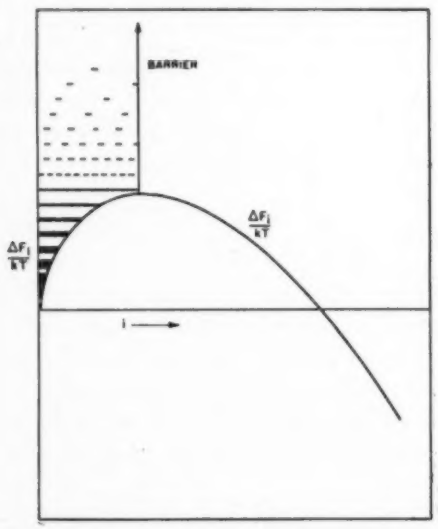


FIG. 4. Equilibrium distribution of embryos for transformation to a more stable phase, assuming a vertical free energy barrier at $(\Delta F_i)_{max}$.

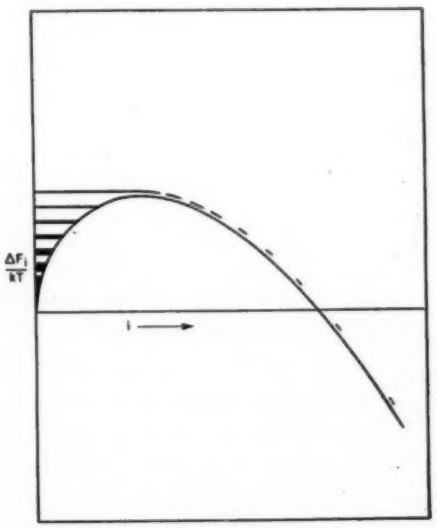


FIG. 5. Steady-state nucleation for transformation to a more stable phase.

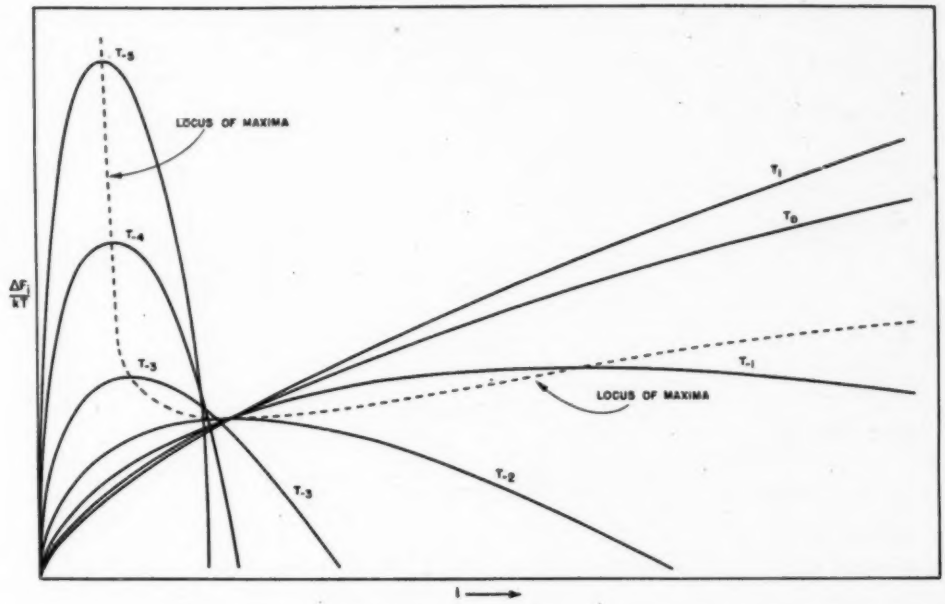


FIG. 6. $\Delta F_i/kT$ *versus* i at several temperatures for a typical phase transformation.

be noted that the steady state rate of nucleation becomes vanishingly small as $\Delta F_i^*/kT$ increases.

The growth of embryos over the free energy hill is termed thermal nucleation. Steady-state thermal nucleation is illustrated in Fig. 5. Transient thermal nucleation occurs whenever the initial distribution of embryos differs from the steady-state distribution; if, for example, there are initially fewer than the steady-state number, an incubation period will be observed. During this period the rate of nucleation will increase to its steady-state value.

ATHERMAL NUCLEATION

Phase transformations are generally produced by cooling or heating a sample of material from a temperature where one phase is stable to a temperature where another phase is stable. If the temperature change is rapid, the initial distribution of embryos at the new temperature will not differ appreciably from that corresponding to the old temperature; there will have been insufficient time for them to shrink or grow to their steady-state concentration. Deviations from the steady-state condition also will be present when the temperature change is gradual. Every phase transformation must, therefore, begin with a transient rate of nucleation.

Consider a transformation from phase α to phase β where α is relatively stable at a high temperature, and β is relatively stable at a low temperature. The equilibrium temperature of the bulk phases in the absence of stress and interfaces is T_0. A family of $\Delta F_i/kT$ *versus* i curves for a typical transformation of this kind is sketched in Fig. 6.

Above the transformation temperature T_0, where phase α is stable, the curve of $\Delta F_i/kT$ *versus* i rises indefinitely and has no maximum. Just below the temperature T_0, $\Delta F_i/kT$ has a maximum at a large value of i. Slightly farther below T_0 the quantity B becomes more negative, moving the peak of the hill downward and to the left. When sufficiently low temperatures are approached, the factor $1/T$ causes the height of the peak to reverse its trend and increase, although the peak may still move to the left if the quantity $(B+C)$ continues to decrease. As $T\rightarrow 0$, the height of the peak increases without limit.

The maxima of the curves of $\Delta F_i/kT$ *versus* i are observed to have a minimum at a temperature below T_0, and to tend to infinity as $T\rightarrow T_0$ or as $T\rightarrow 0$. The steady-state rate of thermal nucleation

$$I_\beta \sim e^{-\Delta F_i^*/kT}$$

is, therefore, a maximum at the temperature

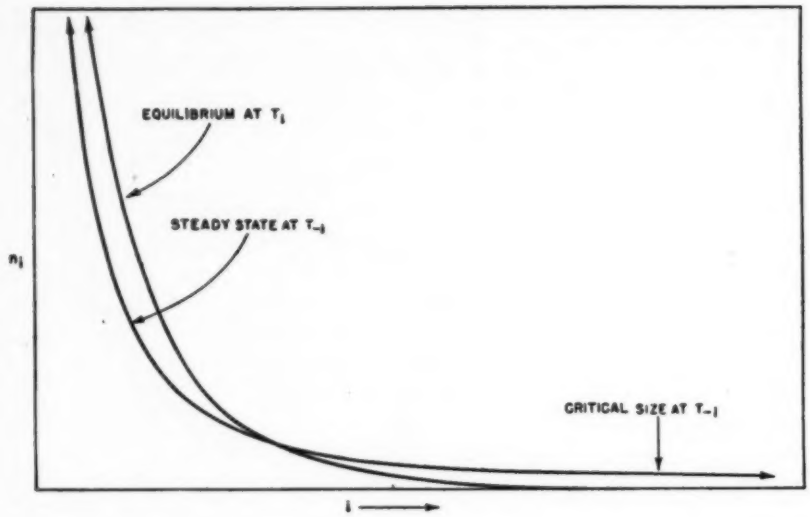


FIG. 7. Distribution of embryo sizes leading to an incubation period.

giving the minimum peak, and becomes zero as the temperature approaches either zero or T_0.

Transformation from phase α to phase β in a sample of the material described in Fig. 6 can be brought about by holding the sample at a temperature above the equilibrium temperature T_0, then cooling it rapidly to a temperature below T_0. The corresponding nucleation process is extremely sensitive to the temperatures and cooling rate of the heat treatment employed. Two illustrative heat treatments will be considered.

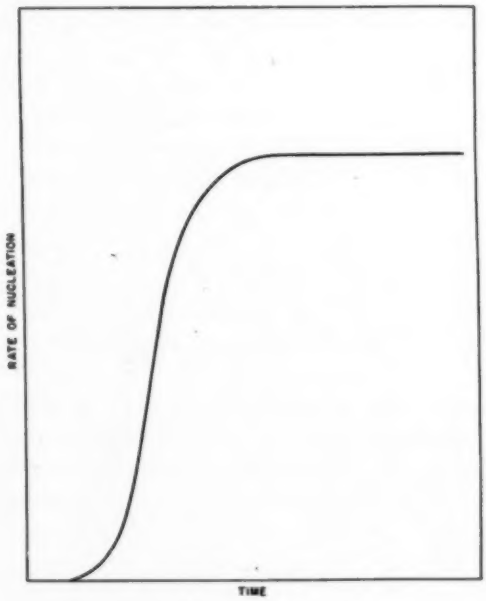


FIG. 8. Incubation period in thermal nucleation.

(1) The specimen is quenched from temperature T_1 above T_0 to temperature T_{-1} somewhat below T_0. The corresponding $\Delta F_i/kT$ *versus* i curves appear in Fig. 6. If the rate of cooling is sufficiently rapid, the high temperature equilibrium distribution of embryos will be retained during quenching. Figure 7 indicates, however, that in this distribution there are no particles exceeding the critical size, although when steady state conditions are realized at the lower temperature, a substantial current of embryos will continue to grow to critical size. Figure 8 gives the curve of rate of nucleation *versus* time for this heat treatment. The time required for the retained embryos to reach critical size by the chain of bimolecular reactions[6] appears as an incubation period.

(2) The specimen is quenched from temperature T_1 to temperature T_{-4} considerably below T_0. The corresponding curves of $\Delta F_i/kT$ *versus* i appear in Fig. 6. Again the high temperature distribution of embryos is retained during quenching. Figure 9 shows that the peak of the low temperature $\Delta F_i/kT$ *versus* i curve falls far to the left, and that a number of particles retained from the high temperature lie to the right of the peak and have become nuclei; these nuclei grow as regions of the new phase β. Similarly, embryos lying to the left of the low temperature peak are present in excess of their steady state concentration, and shrink to the low temperature steady state distribution for which the steady rate of

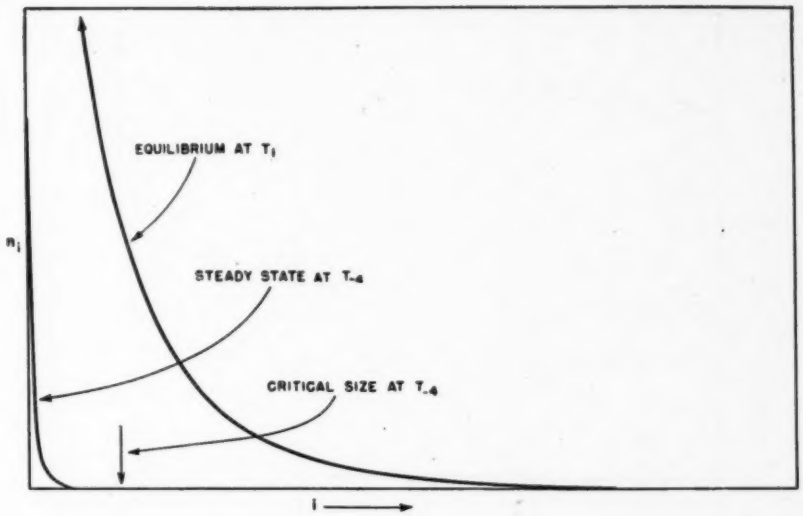


FIG. 9. Distribution of embryo sizes leading to athermal nucleation.

thermal nucleation is effectively zero. The rate of nucleation *versus* time for this heat treatment is shown in Fig. 10.

The nucleation mechanism whereby embryos grow and become nuclei through the action of thermal fluctuations has been termed thermal nucleation. The second nucleation mechanism, whereby embryos are automatically promoted to nuclei as the peak of the free energy hill moves past them to smaller values of i, will be called athermal nucleation.

Many transformations such as austenite→martensite in steels, body centered cubic→face centered cubic in lithium, face centered cubic→hexagonal close packed in cobalt, and $\beta \rightarrow \beta'$ in brass, proceed entirely by the growth of athermal nuclei. For such reactions, nucleation takes place only during cooling, plastic deformation, and other treatments that move the peak of the free energy curve far enough to the left to promote retained embryos to nuclei. Plastic deformation, for example, can move the peak to the

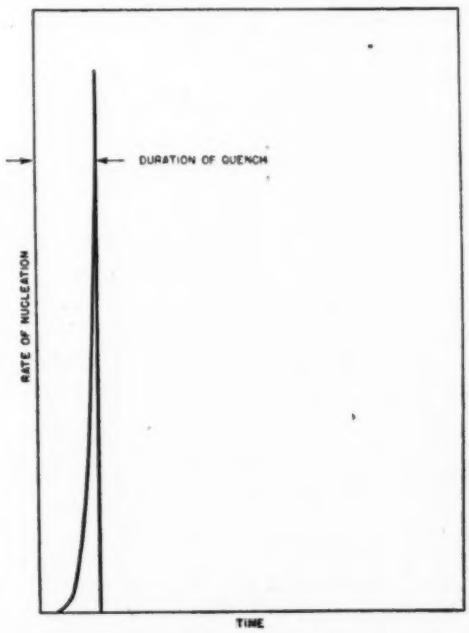


FIG. 10. Athermal nucleation.

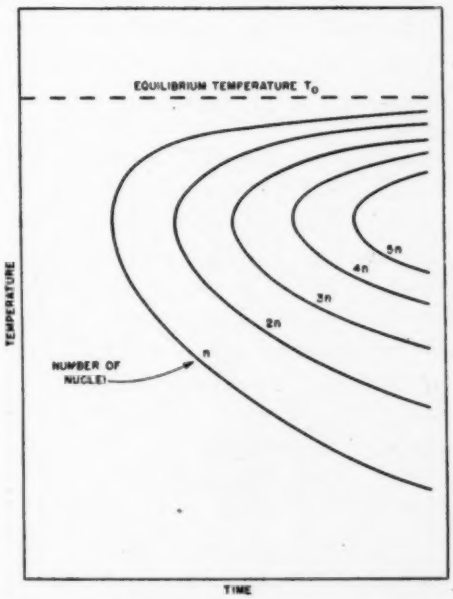


FIG. 11. Time-temperature-nucleation plot for a typical thermal transformation.

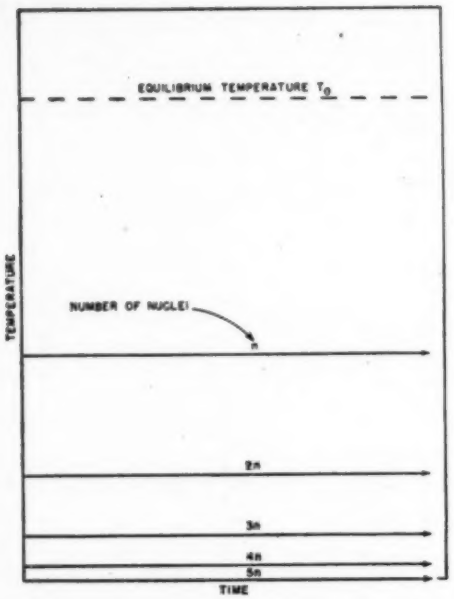


FIG. 12. Time-temperature-nucleation plot for a typical athermal transformation.

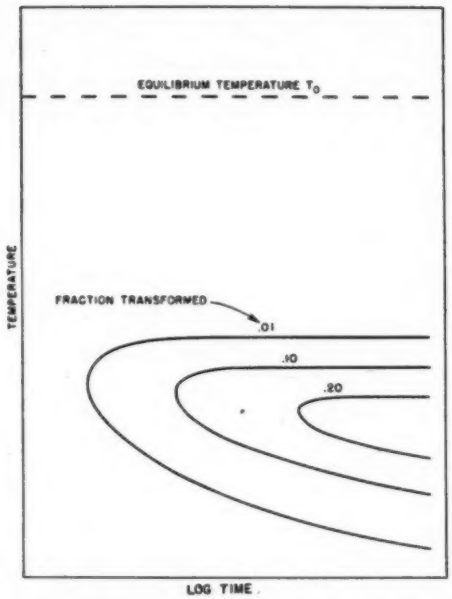


FIG. 14. Time-temperature-transformation plot for a typical athermally nucleated transformation.

left by modifying the strain energy in such a way as to decrease the value of C in Eq. (1). The nucleation mechanism for these transformations, as well as for many more that proceed in part by athermal and in part by thermal nucleation, has not previously been understood.

Thermal and athermal nucleation are contrasted in Figs. 11 and 12. Figure 11 is a time-temperature-nucleation plot for a typical thermal transformation, and Fig. 12 is a similar plot for a typical athermal transformation. A time-temperature-nucleation plot gives the number of nuclei present in the parent matrix as a function of time after instantaneous quenching to an arbitrary temperature T from equilibrium at a

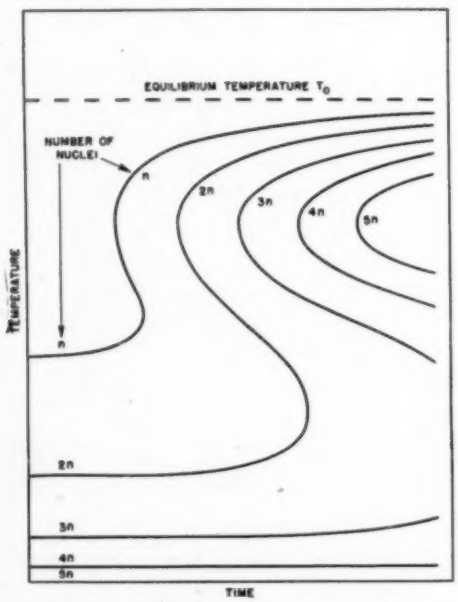


FIG. 13. Generalized time-temperature-nucleation plot.

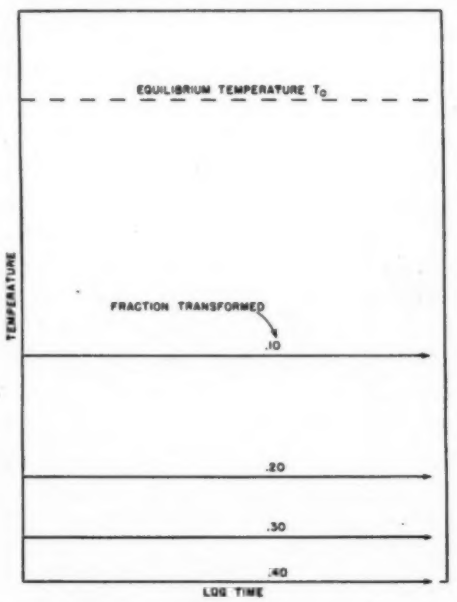


FIG. 15. Time-temperature-transformation plot for a typical martensite-like transformation.

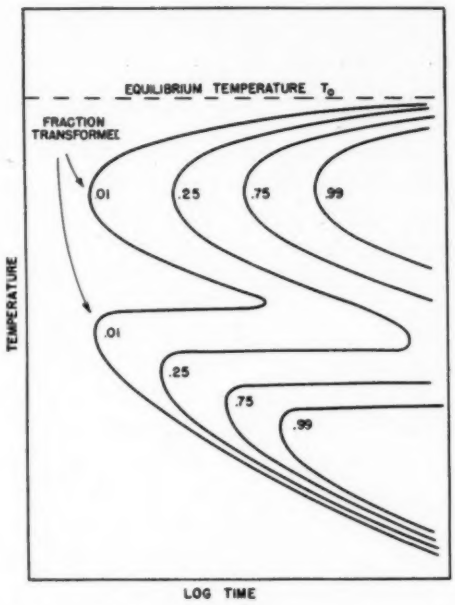


FIG. 16. Generalized time-temperature-transformation plot.

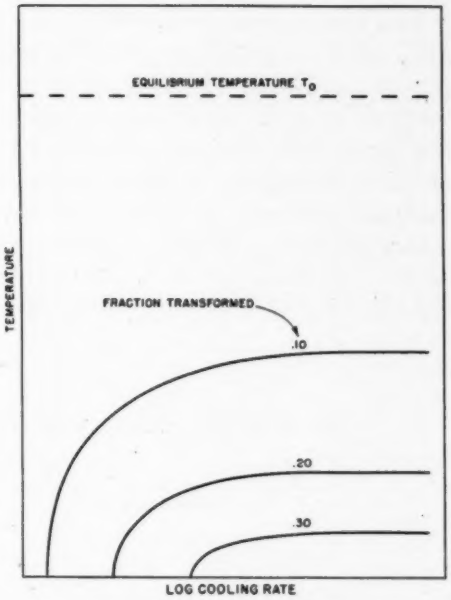


FIG. 17. Stabilization of a typical martensite-like transformation.

fixed temperature T_1 above the transformation temperature T_0. Figures 11 and 12 reveal the primary characteristics of the two types of nucleation. Thermal nuclei are formed at constant temperature and at a rate which is a maximum at a temperature somewhat below T_0. Athermal nuclei appear only while the temperature is changing, and are not formed at constant temperature. Many transformations that proceed by thermal nucleation in one temperature range proceed by athermal nucleation, on cooling, to lower temperatures. The corresponding time-temperature-nucleation plots show a transition from the thermal to the athermal mechanism, as in Fig. 13.

The more familiar time-temperature-transformation plot gives the fraction of transformed material as a function of time after an instantaneous quench to an arbitrary temperature T, from equilibrium at a fixed temperature T_1 above the transformation temperature T_0. The time-temperature-transformation plot of an athermally nucleated transformation resembles that in Fig. 14, for solids where the final size of each transformed region is limited by the grain size of the parent matrix or by included particles of a third phase. The similarity between Fig. 14 and the time-temperature-transformation plots for the transformations austenite→bainite in steels is evident, although after the initial athermal process many bainite reactions proceed slowly by thermal nucleation, causing the lines which are horizontal in Fig. 14 to rise slowly with increasing time.

When the rate of growth of a new phase is very rapid, growth of athermal nuclei is frequently complete before the first observations are made. The experimental time-temperature-transformation relationships for such martensite or martensite-like transformations resemble the plot in Fig. 15.

Figure 16 gives a generalized time-temperature transformation plot for a transformation that proceeds by thermal nucleation in one temperature range and by athermal nucleation on cooling to lower temperatures.

A further characteristic of athermal reactions is the stabilization observed with slower cooling rates. As the cooling rate is decreased for an athermally nucleated process, more and more nuclei have time to shrink ahead of the advancing free energy hill, and the number of athermal nuclei formed in any given temperature interval is decreased. The cooling rate-temperature-transformation plot in Fig. 17 illustrates stabilization of a typical martensite reaction.

Although only interfacial energy, bulk free

energy and strain energy have been considered in detail, other forms of energy are associated with particular transformations. Electromagnetic energy is important in considering the nucleation of magnetic transformations, and strain energy of distortion is important when embryos and nuclei are coherent with the parent matrix.

Further, anisotropy of the elastic constants and of the interfacial energy will modify the shape of particles of a new phase and will determine the relative orientations of new and old phases. Additional forms of energy and the effects of anisotropy can be included in Eq. (1) when required.

The general nucleation theory outlined above is applicable to phenomena other than phase transformations. It applies equally well to the nucleation of fracture cracks and slip bands in solids, and to the nucleation of magnetic transformations.

An Analytical Discussion of the Construction of Pole Figures

JOHN K. WOOD*
Bausch and Lomb Optical Company, Rochester, New York
(Received February 11, 1948)

The construction of pole figures of crystal orientation data obtained by means of x-rays can be made quickly and accurately if the angular relationships between planes or "poles" are known and if all of the data from a diffraction pattern are used. A flexible set of equations is derived from which these relationships can be found starting with customary crystallographic information. Data for an hypothetical case of diffraction are computed to illustrate the method. Although the examples refer, for the most part, to cubic crystals, the method can be extended to cover other types.

INTRODUCTION

IN constructing pole figures[1] of data obtained from x-ray "pinhole" diffraction photographs in crystal orientation studies, a large amount of data is neglected because it does not enter the pole figure construction in a simple manner. The usual procedure is to take the most direct information from a large number of photographs rather than all the data from only a few photographs. Since diffraction photographs require exposures of the order of hours, it would be desirable to get the required information with as few photographs as possible so that the time required to make a complete crystal orientation study could be shortened.

The study of a pole figure involves visualizing a three dimensional picture of a rather complex set of planes and their normals or "poles" and the angular relationships of these poles to the x-ray beam and to the surface of the specimen under scrutiny. However, once a study of this sort has been accomplished, it is only a matter of following rather simple rules to make a pole figure showing the crystal orientation of a given specimen. In this study, a cubic iron crystal will be used as an example.[2] The development, will, however, be general so that it will apply to other crystal forms.

If a crystal is at the center of a sphere of unit radius, the location of any pole of the crystal with respect to a reference pole can be uniquely identified on this sphere by specifying two angles. It is customary in stereographic projection work to use great circles on the projection sphere through the reference or primitive pole as one of these coordinates, defined here as angles of constant δ, and the small circles or latitude circles which have the primitive pole as a center, as the other coordinate defined as constant θ lines. These two sets of mutually orthogonal circles form the coordinate system by which the orientation of any pole may be described. The details of the construction of the stereographic projection

* Work done at the Physics Department, The Pennsylvania State College, State College, Pennsylvania.

[1] See F. Wever, Trans. A.I.M.E. **93**, 51 (1931) for a discussion of pole figures from a different method of approach.

[2] John K. Wood, "Pole Figures of the Effect of Some Cold Rolling Mill Variables on Low Carbon Steel," Trans. A. S. M. **39**, 725 (1947).

of this net need not concern one here since that part of the procedure is taken care of in the "drawing board" technique[3] and is purely mechanical. The present discussion is intended to find what angles are to be plotted. Figure 1 shows a stereographic projection of this network of circles when the reference or primitive pole of the projection sphere is in the plane of projection. If, instead, the primitive pole were normal to the projection plane the constant θ circles would be concentric about the center of the projection circle and the constant δ lines would be radial lines.

The normals or poles to diffracting planes of a crystal act as a rigid system just as do the planes themselves. When one pole is moved about a given point or axis, the others move by the same amount. Thus, when a unit crystal is rotated about some axis through it, the paths of the poles on a stereographic plot will be small circles or latitude lines on the plot, intersecting the axis of rotation at right angles. If the paths were great circles the poles could not act as a rigid system because great circles pass through the same point. Figure 2 shows a pole figure of the {110} and {100} families of a unit cubic structure. If an x-ray beam directed along the Y axis strikes the unit cube while the cube is rotated about the X axis, diffraction will occur whenever a pole subtends the proper angle with the x-ray beam. Figure 3 shows the paths of the {1$\bar{1}$0} family of poles when the unit cube is rotated about the X axis.

Since the plot is usually made in terms of one family to avoid a complicated appearing pole figure, it is important to know where the poles of the plotting family move when a pole from some other family moves to a diffracting position. For example, suppose a pole figure is being plotted in terms of the 101 pole and the data is taken from the diffraction arc made by the 001 pole. Figure 4 shows an hypothetical x-ray photograph of the diffraction due to the 001 pole when the Bragg angle for diffraction is 20°. For simplicity it will be assumed that the x-ray beam is parallel to the reference or projection plane (the specimen plane).

In order to diffract an x-ray beam coming in

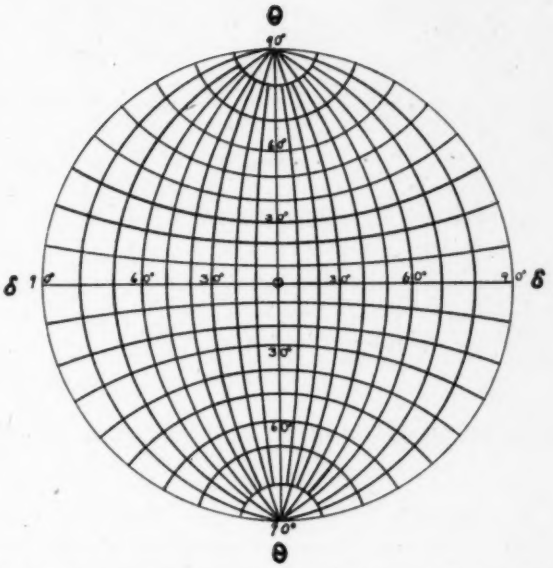


FIG. 1. A stereographic projection of the coordinates. The primitive pole is in the plane of projection.

along the Y axis, the 001 pole must rotate about the X axis, to the great circle $\delta = 20°$ as in Fig. 5. When this happens the 101 pole moves along the small circle $\theta = 45°$ to the great circle $\delta = 20°$ because both poles started from $\delta = 0°$. As the 001 pole then rotates about the Y axis (the x-ray beam) to produce the arc shown in Fig. 4, it moves along the line $\theta_1' = 20°$ while the 101 pole moves along the line $\theta_2' = 14°$ by the same amount.

It is important to note that lines of constant δ are great circles with the X axis as the pole while lines of constant θ' are small circles with the Y axis as the pole. Two sets of coordinates are used. One set, variables δ and θ in Fig. 5, gives the original positions of the poles with respect to some arbitrary reference point and the positions to which they must move in order to be in a diffracting position. The other set, variables δ' and θ', has the x-ray beam as the axis of the projection net and locates the exact diffraction spot or arc. Poles located on the first network must follow small circles on the second network to represent the data of the photograph. This is exactly what is done in an actual situation. The two sets of nets are constructed and used successively to plot the data. Figure 6 shows the arcs on the pole figure described by the {110} family when the 001 pole moves so as to produce the arc on the photograph. This assumes that the

[3] See F. W. Sohon, *The Stereographic Projection* (Chemical Publishing Company, New York, 1941).

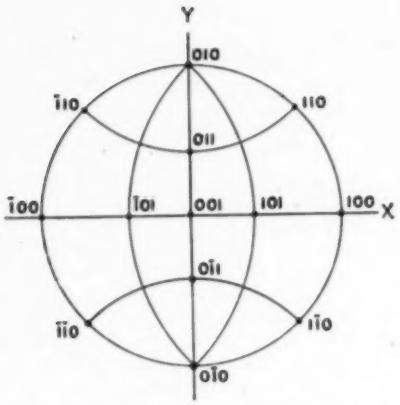


FIG. 2. Pole figure of the {110} and {100} families of a unit cubic crystal.

x-ray beam is in the reference plane which is seldom the case since that plane is usually the surface of the specimen being examined. In practice the x-ray beam makes a finite angle with the surface of the specimen. Then, not the x-ray beam but the projection of the x-ray beam on the reference or plotting plane becomes one of the axes of rotation corresponding to the Y axis in the example just given. The plot of Fig. 6 will be modified when the x-ray beam hits the specimen or reference plane at an angle since the plotting net will be displaced. This effect will be taken into account later.

DEVELOPMENT OF THE EQUATIONS

Three sets of coordinates, whose origins coincide, are used to arrive at a general solution of the problem. The first set is the familiar XYZ coordinates of a unit crystal lattice. The second set contains a quantity of metallurgical importance. In the case of rolled iron it is the rolling direction of the specimen. The rolling direction is called the R axis, the cross rolling direction is the C axis and the normal to the specimen is the N axis. The third set, $X'Y'Z$, has the x-ray beam projected along one of the axes. This projection on either X' or Y' is the pole of the projection sphere.

The direction numbers of the normals or poles of the planes in a crystal are the Miller indices of the associated planes. Since these data are usually given in the XYZ system of coordinates, one must be able to transform these numbers from the XYZ system to any other because the pole figure is usually constructed in one of the other sets of coordinates.

If the reference (XYZ) axes of a line with direction cosines LMN are rotated, the new direction cosines $L'M'N'$ of the line will be the respective sums of the components of LMN along the new axes $X'Y'Z'$. This transformation can be written very concisely in matrix form as:

$$\begin{pmatrix} L' \\ M' \\ N' \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} L \\ M \\ N \end{pmatrix}$$

$$= \begin{pmatrix} Ll_{11}+Ml_{12}+Nl_{13} \\ Ll_{21}+Ml_{22}+Nl_{23} \\ Ll_{31}+Ml_{32}+Nl_{33} \end{pmatrix}$$

where the elements, l_{nm}, of the transformation matrix are unit vectors along the XYZ axes projected on the $X'Y'Z'$ axes. To translate this to an ordinary algebraic expression one would write

$$L' = Ll_{11}+Ml_{12}+Nl_{13}, \text{ etc.}$$

An example of a rotation through an angle δ about the Z axis will illustrate this matrix method of transformation.

$$\begin{pmatrix} L' \\ M' \\ N' \end{pmatrix} = \begin{pmatrix} \cos\delta & \sin\delta & 0 \\ -\sin\delta & \cos\delta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} L \\ M \\ N \end{pmatrix}$$

$$= \begin{pmatrix} L\cos\delta+M\sin\delta \\ -L\sin\delta+M\cos\delta \\ N \end{pmatrix}.$$

Further rotational transformations can be effected by operating on the new set of direction cosines in the same manner. The result is the product of all the transformation matrices operating on the original direction cosines. A second rotation of

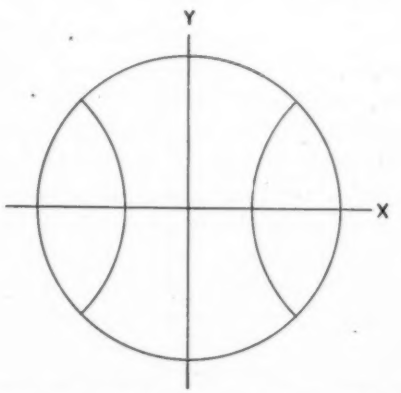


FIG. 3. Stereographic projection of the paths described by the {110} family of poles when the unit crystal is rotated about the X axis.

the above example through an angle θ about the X axis gives

$$\begin{pmatrix} L'' \\ M'' \\ N'' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} L' \\ M' \\ N' \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\delta & \sin\delta & 0 \\ -\sin\delta & \cos\delta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} L \\ M \\ N \end{pmatrix}$$

where L, M, N are the direction cosines after two transformations.

The next step is to find the position of any pole of a diffracting crystal in terms of the direction cosines of the pole, the angle between the $X'Y'Z'$ system of coordinates, which has the projection of the x-ray beam along one axis, and the RCN system which has a metallurgical quantity along one axis. The angle between the XYZ reference coordinates and the $X'Y'Z'$ coordinates must be known to determine the direction cosines of the pole in the $X'Y'Z'$ coordinates.

The axis of the stereographic network, which is the X' axis of Fig. 7, makes an angle z with the rolling direction of the specimen. The meanings of the variables are as follows:

z = angle between the rolling direction and the X' axis;

A = angle between the Y' axis and the trigonometric projection of the pole on the $X'Y'$ plane;

φ = angle between the RN plane and the plane through the pole and the N axis. It is equal to $90° - z + A$;

K = angle between N (or Z') and the pole;

θ = small circle coordinate used in the stereographic projection of the pole (Note that the primitive pole is the Y' axis);

δ = great circle coordinate of the pole used in the stereographic projection;

α = angle between the x-ray beam and the RC plane.

Let L, M and N now be the direction cosines in the RCN coordinates. From Fig. 7,

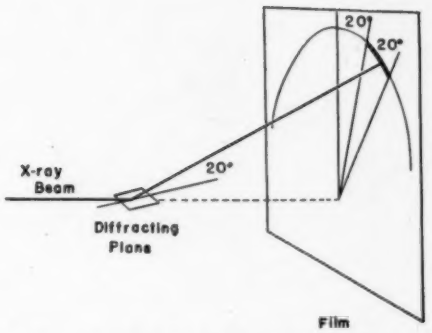


FIG. 4. Hypothetical situation of diffraction due to the 001 pole when the Bragg angle for diffraction is 20°. Wide arc indicates the most probable position.

$$\sin\theta = \sin K \cos A, \tag{1}$$

$$\sin\delta = \frac{\sin K \sin A}{\cos\theta}. \tag{2}$$

From Eq. (1),

$$\cos\theta = (1 - \sin^2 K \cos^2 A)^{\frac{1}{2}}$$

which is substituted into Eq. (2), giving

$$\sin\delta = \frac{\sin K \sin A}{(1 - \sin^2 K \cos^2 A)^{\frac{1}{2}}}. \tag{3}$$

To solve Eqs. (1) and (3) in terms of L, M, N and z, the following equations are needed:

$$\cos A = \sin(\varphi + z) = \sin\varphi \cos z + \cos\varphi \sin z, \tag{4}$$

$$\sin A = -\cos(\varphi + z) = \cos\varphi \cos z - \sin\varphi \sin z, \tag{5}$$

$$\sin K = (L^2 + M^2)^{\frac{1}{2}}, \tag{6}$$

$$\sin\varphi = \frac{L}{(L^2 + M^2)^{\frac{1}{2}}}, \tag{7}$$

$$\cos\varphi = \frac{M}{(L^2 + M^2)^{\frac{1}{2}}}. \tag{8}$$

Substituting Eq. (4) and then Eqs. (6), (7) and (8) in Eq. (1) gives

$$\sin\theta = L \cos z + M \sin z \tag{9}$$

and substituting Eqs. (4) and (5) and then Eqs. (6), (7) and (8) in Eq. (3) gives

$$\sin\delta = \frac{L \sin z - M \cos z}{(1 - L^2 \cos^2 z - M^2 \sin^2 z - 2ML \cos z \sin z)^{\frac{1}{2}}}. \tag{10}$$

From the L, M, and N values computed above, the values of θ and δ can be computed for the various values of z.

If the coordinates of the diffracting position can be obtained for any pole, the rotation of the crystal from some arbitrary or ideal position can

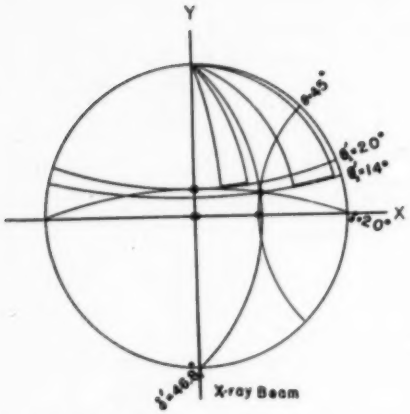


FIG. 5. Projection of the movement of the 101 pole when the 001 pole caused the diffraction shown in Fig. 4.

□—001 pole in original position.
■—001 pole in a diffracting position.
○—101 pole in original position.
●—101 pole when the 001 pole is in a diffracting position.

be found and since the plotting poles will rotate the same amount from that arbitrary position, the pole figure can be constructed from this information.

The direction cosines of the x-ray beam and of the pole can be found in the same set of coordinates, so it is a simple matter to find the angle between them which would be equal to the diffracting angle for that particular pole. If the beam is in the $X'Z'$ plane and makes an angle α with the specimen, its direction numbers in the RCN system are,

$$L_1=\cos\alpha\,\cos z,$$
$$M_1=-\cos\alpha\,\sin z,$$
$$N_1=\sin\alpha.$$

The direction cosines of the pole are most easily found in the $X'Y'Z'$ system and are,

$$L_2'=\cos\theta\,\sin\delta,$$
$$M_2'=\sin\theta,$$
$$N_2'=\cos\theta\,\cos\delta.$$

These direction cosines are transformed to the RCN system by the method outlined previously. The new direction cosines of the pole are,

$$L_2=\sin\theta\,\sin z+\cos\theta\,\sin\delta\,\cos z,$$
$$M_2=\sin\theta\,\cos z-\cos\theta\,\sin\delta\,\sin z, \qquad (11)$$
$$N_2=\cos\theta\,\cos\delta.$$

The expression for the angle between the two lines will then be,

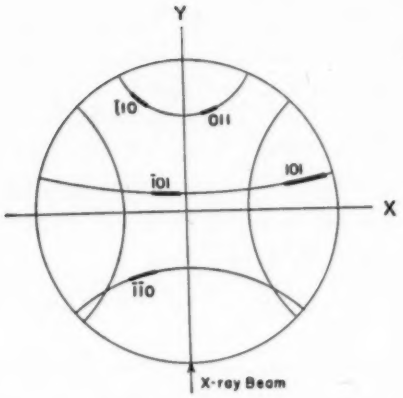


FIG. 6. Pole figure of the arcs described by the {110} family of poles when the 001 pole moves to produce the arc in Fig. 4.

$$\cos(90^\circ-D)=\sin D=L_1L_2+M_1M_2+N_1N_2$$
$$=\cos\alpha\,\cos z(\cos\theta\,\sin\delta_1\,\cos z+\sin\theta\,\sin z)$$
$$+\cos\alpha\,\sin z(\cos\theta\,\sin\delta_1\,\sin z-\sin\theta\,\cos z)$$
$$+\sin\alpha\,\cos\theta\,\cos\delta_1$$

where δ_1 is now the diffracting position of the pole and D is the diffracting angle of the pole in question. Solving this for δ_1

$$\sin(\delta_1+\alpha)=\frac{\sin D}{\cos\theta}. \qquad (12)$$

The rotation of the crystal about an axis at right angles to the beam is equal to the difference between this value of δ_1 and the initial δ value from Eq. (10).

DISCUSSION

The mathematical discussion will become more clear if the actual angular positions, plotted in Fig. 5 from the data of Fig. 4, are computed. It

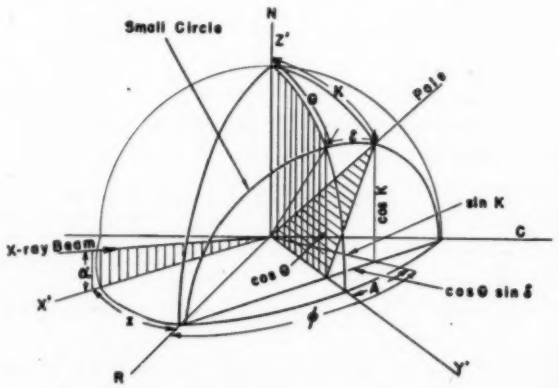


FIG. 7. Path of a pole when its corresponding plane rotates about an axis at Y'.

will be recalled in the discussion of Fig. 4, that the Bragg diffraction angle is 20° and that the x-ray beam is in the plane of the specimen making $\alpha = 0°$. It is assumed, for simplicity, that the three sets of coordinates, XYZ, $X'Y'Z'$ and RCN, coincide. The direction cosines for the 001 pole are $0, 0, 1$ and for the 101 pole are $1/\sqrt{2}, 0, 1/\sqrt{2}$ in the XYZ system. Substituting these values into Eqs. (9) and (10), one gets for the 001 pole $\theta = 0°$, $\delta = 0°$, and for the 101 pole $\theta = 45°$, $\delta = 0°$. These values give the original positions of the poles on the stereographic net and can be plotted. Now if the 001 pole is to cause diffraction of the x-ray beam, it must move to $\delta = 20°$ which is computed from Eq. (12). Since the system of poles is rigid, the 101 pole must move the same amount to the position $\theta = 45°$, $\delta = 20°$. The new positions can now be plotted. The 001 pole is now in a position to diffract but it must rotate about the beam or Y axis until it is in the exact spot or arc indicated by the data on the photographic film. The position of the pole must now be found in coordinates whose axes are at right angles to the first set so that the small circle which the pole follows as it rotates about the beam may be determined.

The direction cosines of the pole in the new position are found by substituting $\theta = 45°$, $\delta = 20°$ in Eq. (11). These direction cosines are substituted in Eqs. (9) and (10) making $z = 90°$ so that the new net will be at right angles to the first net. From these values, $\theta' = 14°$, $\delta' = 46.8°$, one can plot the circle on which the 101 pole moves when diffraction takes place. It is now a matter of marking off the limits of the arc on the small circle $\theta' = 14°$ to complete the pole figure.

The assumptions in this discussion were made to simplify the presentation. Actually, it is impossible for the x-ray beam to be in the plane of the specimen because it would be very difficult to get a diffraction photograph. The stereographic net mentioned in the example would not look like Fig. 1 as was assumed but would look like Fig. 8, where the x-ray beam or primitive pole of the projection sphere is not in the projection plane but makes an angle of 10° with the projection plane. The computations would be exactly the same with $\alpha = 10°$ in the equations but the drawing board work would be different. It was assumed that the three coordinate systems were the

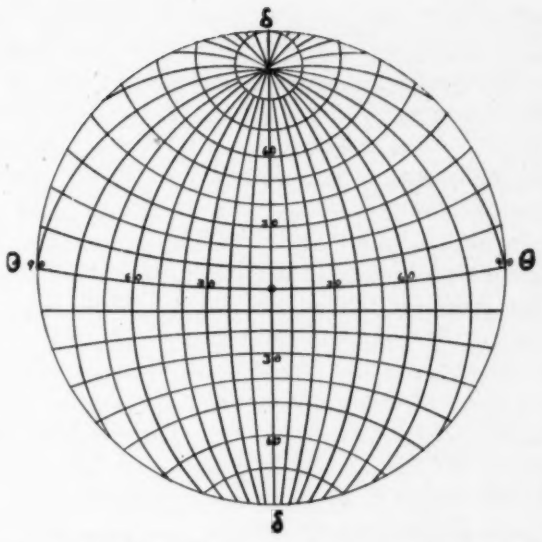


FIG. 8. Stereographic net showing lines making a constant angle with the x-ray beam. The x-ray beam makes an angle of 10° with the plane of the specimen.

same. In a general case, the angle z between the rolling direction and the x-ray beam is determined from the physical arrangement of the experiment. This is the angle between the $X'Y'Z'$ and RCN systems. The angle between the XYZ and RCN sets of coordinates can be obtained from a transmission diffraction photograph where the beam is normal to the specimen. In a specimen with oriented crystals, this is the angle between the "most probable" positions of poles and the rolling direction. The rolling axis is known from the experiment.

Since it is now possible to use all of the available data from one diffraction photograph, it is important to see how many photographs are

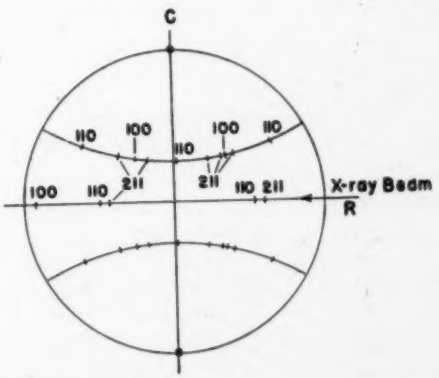


FIG. 9. Positions of the {110} family when any plane from the {110}, {100} or {211} families is in a diffracting position for $z = 0°$, $\alpha = 10°$.

needed for a single pole figure. In the case of rolled sheet iron, the data from the first three arcs on the diffraction photograph taken with the x-ray beam at 10° from the specimen plane and parallel to the rolling direction results in fifty plotting points. These plotting points are positions of the {110} family when the {110}, {100} and {211} families are in diffracting positions. Half of these points are shown in Fig. 9 while the other half are found by rotating this figure through 180°. If the x-ray beam is 10° from the specimen plane and parallel to the cross-rolling direction, another fifty plotting points are located. These are found by rotating Fig. 9 through 90° for twenty-five points and through 270° for another twenty-five. A normal transmission diffraction photograph will yield another twelve plotting points as well as the angle between the *XYZ* and *RCN* systems as mentioned above. This photograph also provides data for the periphery of the pole figure while the two reflection photographs yield information for the central portion of the pole figure.

A pole figure can be made rapidly from x-ray diffraction data by taking a drawing such as Fig. 9 with the positions of the plotting family when various poles are in diffracting positions and placing it on a net such as Fig. 8 which shows the paths these plotting poles will follow to coincide with the data on the photograph. A sheet of paper is placed on the superimposed drawings and the whole is placed on an illuminated viewing box. The positions of the plotting poles can then be marked on the sheet of paper from the data on the photograph. The three photographs mentioned, a normal transmission diffraction pattern, a reflection pattern parallel to the rolling direction, and a reflection pattern perpendicular to the rolling direction will yield enough data for a well defined pole figure.

In plotting the data on a pole figure the relationship between the poles of the single crystal and the diffraction photograph must be kept constantly in mind. In many cases, if certain poles are to diffract, others cannot because they would be too far from a diffracting position. In particular cases, this dilemma can be resolved from a qualitative study of the physical situation. Thus, one may find arcs on the photograph which are impossible if the crystal is required to be in only one position because they were made from crystals in two or more positions. One pole could diffract while the other could not in each position. The intensity ratio of more intense arcs to less intense arcs aids in interpreting these cases. It is this sort of situation that one wishes to show in a pole figure.

The advantage of this method is that it gives quantitatively and quickly the contributions to a pole figure of a large number of poles. A mechanical projecting device such as the McLachlan projection machine[4] can be used to plot the data using the angles obtained from Eqs. (10) and (12) as the constants of the machine, or they can be plotted by hand using the drawing instruments methods of stereographic projection.

[4] D. McLachlan, Rev. Sci. Inst. **7**, 301 (1936).

An Electron Microscope Study of Gold Smoke Deposits*,**

LOUIS HARRIS AND DAVID JEFFRIES
Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

BENJAMIN M. SIEGEL***
Weizmann Institute of Science, Rehovoth, Palestine

(Received February 18, 1948)

Gold smoke deposits have been examined at high resolution with the electron microscope, and the size and distribution of the unit particles determined. The smokes were produced by evaporating gold from a hot tungsten filament in a partial atmosphere of nitrogen, and the samples were collected on thin collodion films. It was found that the size of the unit colloidal gold particles varied with the pressure of the atmosphere and the rate of evaporation. The manner of aggregation of the particles was found to be independent of the nitrogen atmosphere pressure and rate of evaporation over a relatively wide range. The aggregations, however, changed markedly when a small amount of oxygen was present during the evaporation. Measurements in the visible and infra-red regions of the spectrum were correlated with the electron microscope investigations, and it was found that the optical properties of the deposits depend strikingly upon the manner in which the smoke particles aggregate and how the aggregates are distributed.

INTRODUCTION

METAL and metal oxide smokes have been examined with the electron microscope,[1] but apparently no attempt has been made to study the factors determining the nature of the smoke deposits or to correlate their appearance in the electron microscope with other physical properties. In this investigation we have obtained gold smokes with reproducible optical properties and have examined their particle size and manner of aggregation. The factors controlling the nature of the deposits were varied, and the effects of the variation on particle size, aggregation of particles, and optical properties, were observed. In this paper we shall confine ourselves to the presentation of the results of the electron microscope investigation, including only a brief correlation with the spectral transmission measurements in the visible and infra-red region. The details of the complete investigation of the optical properties have been reported elsewhere.[2]

* This work has been supported by the Navy Department, Bureau of Ships, under Contract 25391.

** Contribution from the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts.

*** This author was at the Polytechnic Institute of Brooklyn by special arrangement during the period in which this work was conducted.

[1] H. Mahl, Jahrb, du A. E. G. Forschung **7**, 67 (1940); R. B. Barnes and C. J. Burton, Ind. Eng. Chem. News **19**, 965 (1944); J. Gundermann and H. Kübz, Kolloid Zeits. **98**, 287 (1942).

[2] L. Harris and R. T. McGinnies, J. Opt. Soc. Am. **38**, 582 (1948).

EXPERIMENTAL PROCEDURE

The smokes were obtained by evaporating the gold from a hot tungsten filament in a partial atmosphere of nitrogen. They were deposited on collodion membranes about 500A thick stretched over a three-centimeter diameter hole in a brass frame. It was necessary to place a metal backing behind the frame, and within 0.1 cm of the film, in order to obtain a sufficiently heavy deposit on the film.

Before setting the film containing the deposit on the copper Electromesh screen for the electron microscope investigation, the transmission of the deposit to the radiation from a 50°C blackbody was measured. As has been shown,[2] this provides a good measure of the mass of the deposit if the conditions under which the deposit was produced are known.

The electron microscope used in this investigation was an RCA type EMU. A self-biased saturated emission gun was used. The intense electron beam introduces the possibility of a change of structure of the deposit under bombardment. To minimize the possibility of such an effect, the micrographs were taken from an area near the center of the open part of the screen, and this area was subjected only to illumination by a beam far from the "cross-over" point. There was no apparent effect from the electron beam. The objective pole piece of the microscope was corrected for asymmetry by the

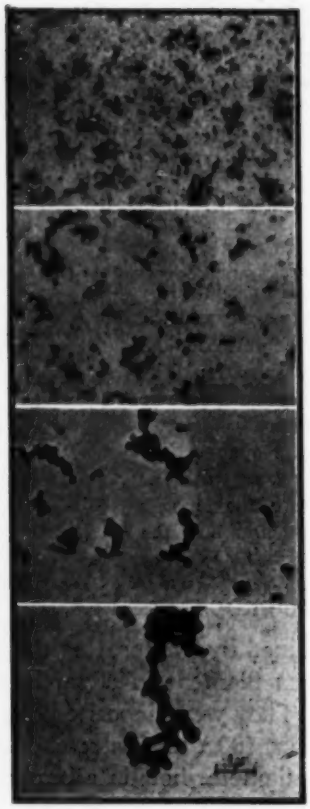


FIG. 1. Extremely thin deposits of gold smokes produced by evaporation at a rate of 3 mg per second but at different pressures of nitrogen: from top to bottom: 0.3 mm, 0.5 mm, 1.0 mm, and 3.0 mm of Hg.

method described by Hillier and Ramberg,[3] and an auxiliary optical microscope viewer giving an additional 10× magnification was used in focusing. The instrument proved to have high resolving power after taking the suggested precautions; clearly defined pictures of particles which could be no greater than 10A in diameter have been obtained. The substrates were slightly thicker than the ones customarily used in electron microscopy, but the high density of the specimen gave good contrast nevertheless. No limiting aperture was used in the objective pole piece.

OBSERVATIONS AND RESULTS

In this investigation the parameters which were varied were the pressure of the nitrogen atmosphere and the rate of evaporation of the gold from the filament. The effects of small amounts of oxygen were also observed.

[3] J. Hillier and E. G. Ramberg, J. App. Phys. **18**, 48 (1947).

(a) Variation of the Pressure of the Nitrogen Atmosphere

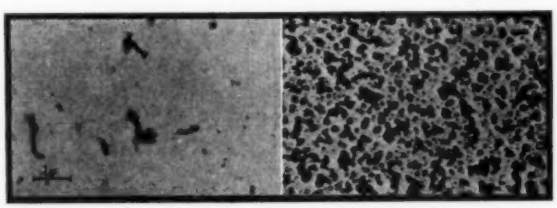
Figure 1 shows at high magnification a series of extremely thin deposits made at pressures of nitrogen ranging from 0.3 mm to 3.0 mm of Hg; the rate of evaporation was kept constant at 3 mg/sec. The variation of particle size with pressure is clearly visible. The lowest pressure produces a deposit of small colloidal particles ranging in diameter from 15A to 100A, and having an average diameter of about 40A. Shadow cast specimens show that the particles are spheroidal. The deposit made in 0.5 mm of nitrogen has larger particles averaging 75A in diameter. The deposits made in 1.0 mm and 3.0 mm of nitrogen show very few individual particles, but the aggregates appear to be made up of individual particles having diameters of about 125A and 200A, respectively.

(b) Variation of Rate of Evaporation

Figure 2 depicts some thin deposits made at very slow and very fast rates of evaporation. The nitrogen pressure in all cases was 0.5 mm. The slow rate produces clusters which have very small unit colloidal particles, while the fast rate produces a deposit of particles having a wide range of size and covering the area uniformly.

(c) Aggregation and Distribution in Deposits

When larger amounts of gold smoke were deposited on collodion films, the layers built up in the same manner for a wide range of nitrogen pressure and rate of evaporation, regardless of the variation in particle size. Figure 3 is typical of a deposit of 15×10^{-6} g/cm² of gold produced in 1.0 mm of nitrogen and at a rate of evaporation of 3 mg/sec. It shows the spongy appearance and the spacing between dense areas characteristic



(a) (b)

FIG. 2. Thin deposits of gold smoke produced in 0.5 mm of nitrogen at slow and fast rates of evaporation: (a) 0.5 mg/sec., (b) 15 mg/sec.

FIG. 3. Gold smoke deposited in 1.0 mm of nitrogen at an evaporation rate of 3 mg/sec.

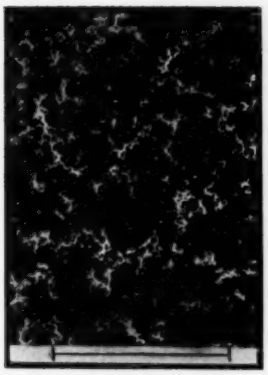


FIG. 4. Gold smoke deposited in 0.5 mm of nitrogen at an evaporation rate of 15 mg/sec.

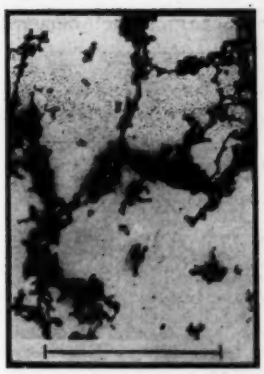


FIG. 5. Gold smoke evaporated at a rate of 3 mg/sec. in an atmosphere of 1.32 mm nitrogen contaminated by 0.08 mm oxygen.

of all heavy deposits produced in 0.5 mm to 3.0 mm at this rate.

The type of heavy deposit obtained by faster evaporation of gold, or in an atmosphere of lower partial pressure of nitrogen than the limits given above, is shown in Fig. 4. This deposit was produced in 0.5 mm of nitrogen and at the high rate of 15 mg/sec. A deposit of similar appearance is obtained by evaporating gold in 0.3 mm of nitrogen at 3 mg/sec.

(d) Effect of Oxygen

A small amount of oxygen in the bell jar during the evaporation causes a marked change in the manner of aggregation and the distribution of the aggregates. Figure 5 is an electron micrograph of a deposit of 15×10^{-6} g/cm² of gold produced in 1.32 mm of nitrogen and 0.08 mm of oxygen, the rate of evaporation being 3 mg/sec. An open irregular network is formed in which the gold is massed densely in a chain-like structure, leaving large areas open. Table I shows clearly how the type of deposit and the infra-red transmission depend upon the amount of oxygen present.

(e) Electron Diffraction Observations

The gold smokes produced in the range of conditions given above were all examined by electron diffraction. None of the deposits gave any evidence of orientation, and only the lattice spacing of pure gold was apparent. The presence of oxygen during the evaporation produced no observable change in the electron diffraction pattern of the deposit.

To determine how oxygen might affect the deposition of gold, collodion films were exposed to the incandescent tungsten filament without gold in exactly the same manner as that used to produce the smoke deposits. When pure nitrogen was present in the bell jar, the electron diffraction pattern of the exposed film showed only the same diffuse rings given by a blank collodion film. When oxygen was present, however, in only the small quantity required to affect the optical properties of a gold smoke, there was a strong diffuse background over the electron diffraction pattern, indicating the presence of an amorphous deposit on the film.

The following experiments indicate that the amorphous layer is an oxide of tungsten. Measurement of loss in weight of the tungsten filament (in absence of gold), correlated with the amount of oxygen consumed, leads to the conclusion that tungsten trioxide, WO_3, is deposited when the ni-

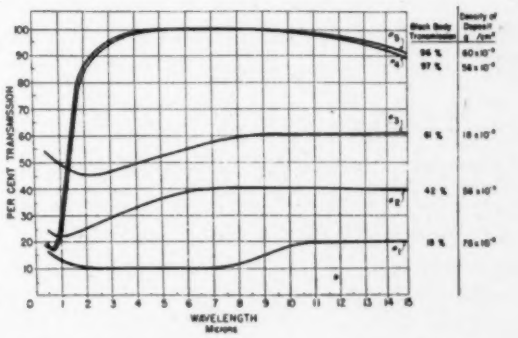


FIG. 6. Spectral transmission of gold smokes.

TABLE I. Samples weighing approximately 15×10^{-6} g/cm².

No.	Pressure (mm) Nitrogen	Oxygen	Rate of evap. mg/sec.	Nature of deposit	50°C blackbody infra-red transmission (%)
6	1.0	.00	.5	spongy, fairly uniform, similar to Fig. 3	65
7	1.32	.08	.5	chain network; broader and denser masses than in Fig. 5	98.5
8	1.0	.00	3	spongy, fairly uniform, Fig. 3	66
9	1.32	.08	3	chain network, Fig. 5	98.5
10	3.00	.009	3	spongy, fairly uniform, similar to Fig. 3	69
11	3.32	.08	3	larger particles forming open network similar to Fig. 5	99
12	3.00	.009	.5	chain network, but broader, more widely spaced chains than in Fig. 5	98.5
13	3.32	.08	.5	chain network; very broad and open-spaced, similar to Fig. 5	99

trogen contains oxygen. We estimate that the tungsten oxide may comprise as much as 10 percent of the weight of the gold smoke, when the partial pressure of oxygen is 0.08 mm in the scattering gas. Less than 0.5 percent by weight of tungsten oxide is present in deposits formed when the partial pressure of oxygen is less than 0.003 mm. Spectroscopic analysis showed conclusively that tungsten was present in gold smokes prepared in an atmosphere of 1 mm of nitrogen and 0.08 mm of oxygen, while the tungsten lines were not observed in smokes prepared in an atmosphere free of oxygen.

(f) Optical Characteristics

Observations with the electron microscope have been correlated with visible and infra-red transmission measurements. Figure 6 shows typical spectral transmission from 0.5–15 microns of gold smokes made under various conditions. Curves No. 1 and No. 3 represent deposits made in 1.0 mm of nitrogen and at a rate of evaporation of 3 mg/sec, and are typical of deposits in 0.5 mm to 3.0 mm of nitrogen and rates of 0.5–3.0 mg/sec. They show the lowest transmission and may be used as excellent infra-red absorbers of very low thermal mass. The deposits produced at lower pressures or higher rates are not as good infra-red absorbers and have characteristics like curve No. 2. The layers having high transmission in the infra-red and high absorption at wave-lengths shorter than 1 micron (curves No. 4 and No. 5) are produced when a small amount of oxygen (0.01 mm to 0.08 mm) is present. Slower rates or higher pressures were found to be best for this type of deposit. The sharp cut-off and high transmission at wave-lengths greater than 3 microns are useful filter characteristics.

DISCUSSION OF RESULTS

The joint electron microscope and optical studies of gold smoke deposits indicate that the manner in which the aggregates build up determines the optical properties. One might expect that since the particle size is in every case very small compared to the wave-length of the incident infra-red radiation, variation in particle size should not be the primary factor determining absorption in this region. Thus it has been found that both the smallest and the largest particles examined can produce either spongy deposits, corresponding to high infra-red absorption, or open chain-like networks, corresponding to low absorption.

It was found that spongy layers having high infra-red absorption could be produced only in pure nitrogen. The presence of a small amount of oxygen results in open-network deposits having very much less absorption beyond 1 micron, while retaining high absorption in the visible spectrum. This phenomenon appears to be due to the simultaneous deposition in the presence of oxygen of an amorphous layer of tungsten oxide which causes an entirely different manner of aggregation of the gold particles.

Letters to the Editor

Prebunching: A Method Controlling Synchrotron Radial Oscillation Amplitudes*

D. C. DEPACKH AND M. BIRNBAUM
Naval Research Laboratory, Washington, D. C.
June 7, 1948

IT has not been generally appreciated that the radial oscillations in synchrotron accelerators, occasioned by the transition from betatron injection, can be minimized by applying the radiofrequency accelerating field considerably before the failure of the betatron flux condition and in accordance with a definite law of increase. This process can be carried on when the electrons have attained nearly their ultimate velocity, so that little or no frequency modulation is required. The technique results in a substantial reduction of the required beam aperture and a consequent reduction in stored energy.

If the r-f is applied in accordance with a power law such that the ultimate voltage, V_0, is attained in a time τ and therefore that the r-f envelope is given by $V_0(t/\tau)^s$, then, as has been shown by Saxon and Schwinger, the electron phase, ϕ, with respect to the rotating r-f phase front satisfies the equation

$$d/dt(H\dot{\phi})+a^2V_0(t/\tau)^s\sin\phi=0, \qquad (1)$$

provided that the betatron flux condition is satisfied over this interval. Here a^2 is a constant relating to the dimensions and geometry of the machine and H is the orbital magnetic field. The same writers have shown that the radial oscillation amplitude x corresponding to the phase displacement ϕ for an electron traveling at nearly its final velocity is given by

$$x=\dot{\phi}R^2/c. \qquad (2)$$

It has been found that there is a broad optimum value for the parameter s which minimizes the peak value of x. This best exponent lies in the range of 4–12 for most practical cases. We have also learned that the oscillation amplitudes are reduced approximately as τ^{-1} provided s has its optimum value. This dependence has been established theoretically and by a numerical solution of (1). A paper containing the analytical solution and a more general treatment of the problem is in preparation.

For a hypothetical accelerator whose radius is 3 meters and in which the transition from betatron to synchrotron operation occurs at 5 Mev, the oscillation amplitudes are given in Table I for a bunching time $\tau=100$ μsec. The magnetic field is taken to increase from 14 to 56 gauss during this time, while the r-f voltage rises from zero to 135,000 volts, which is 10 times the energy gain of the electrons per turn, assuming that they gain 600 Mev in one quarter-cycle of 60 c.p.s. alternating current. The entry $s=0$ corresponds to the case of sudden application of the r-f at a field strength of 56 gauss. These amplitudes were obtained by joining a numerical solution of the bunching equation near the origin to the adiabatic solution at a point at which the steady state has been firmly established. The numerical solution of the bunching equation required for the values in Table I was performed with the aid of a Bell Telephone Laboratories Relay Digital Computer. The writers acknowledge the assistance of Mr. C. H. Chrisman in this connection.

TABLE I.

s	x_{peak} cm
0	76.8
0.5	49.8
1	22.5
2	10.9
3	9.15
4	8.34
6	8.22
8	8.25
12	8.79

* A portion of the work herein reported was submitted by one of the authors (M. Birnbaum) in partial fulfillment of the requirements for the degree of Master of Science at the University of Maryland.

Limiting Sensitivity of a Microwave Spectrometer*

C. H. TOWNES AND S. GESCHWIND
Columbia University, New York, New York
May 13, 1948

IN discussing the minimum detectable absorption in microwave spectroscopy, Hershberger[1] has assumed that a variation in power level at the receiver due to absorption by a gas can be detected if it is comparable with $4kTN\Delta f$. Here k and T have their usual meaning, N and Δf are respectively the noise figure and band width of the receiver. It appears to the authors that this assumption is incorrect, and that instead the detectable variation in power must be comparable to $(kTN\Delta f/m)$, where m is approximately the fraction of power absorbed by the gas, or the modulation coefficient. The correct expressions for optimum length and minimum detectable absorption coefficient for a microwave spectrometer using a wave guide as an absorption cell as thus modified become

$$l_{op}=2\alpha_c^{-1} \qquad (1)$$

and

$$\alpha_{gmin}=2e\alpha_c[2kTN\Delta f/P_0]^{\frac{1}{2}}, \qquad (2)$$

where l_{op} = optimum length of absorption cell, α_c = attenuation coefficient (in cm^{-1}) of wave guide, α_{gmin} = minimum detectable absorption coefficient of gas, P_0 = power introduced into the wave guide before absorption, and e = base of natural logarithms.

For convenience, and to put our expressions into a form familiar to lower frequency theory, let the microwave be represented by a voltage defined so that $V^2=2ZP$ where P is the power and Z the guide impedance. Assume an idealized spectrometer for which the microwave intensity introduced into the absorption cell is of constant frequency $\omega/2\pi$ and the transmitted voltage of the form $V\cos\omega t(1-m\cos\omega_0 t)$. $Vm\cos\omega t\cos\omega_0 t$ represents a variation in the transmitted intensity due to a periodic intro-

duction and withdrawal of gas from the absorption cell at a frequency $(\omega_0/2\pi)$. All the information due to the gas absorption is in two side bands of frequency $\omega+\omega_0$ and $\omega-\omega_0$ which contain power approximately $m^2V^2/4Z$, and it is this sideband power which must be larger than or comparable to the noise $kTN\Delta f$ at the sideband frequencies in order to be observable. The envelope of power transmitted through the guide will be $(V^2/2Z)(1-2m\cos\omega_0 t+m^2\cos^2\omega_0 t)$, where the second term $m\cos\omega_0 t(V^2/Z)$ may be regarded as the power variation due to gas absorption. Hershberger's assumption is equivalent to letting $mV^2/4Z$ be comparable with $kTN\Delta f$ for minimum detectable signal, and since m is a very small number $(\sim10^{-6})$, this differs radically from comparing the sideband power, $m^2V^2/4Z$ with $kTN\Delta f$.

An alternate approach is to discuss the variation in the power envelope produced by noise. Since the signal variation is of the form $Vm\cos\omega t\cos\omega_0 t$, we must look for noise of the same form which can therefore not be distinguished from the signal by any type of detecting circuit. We select noise terms in a small range of frequencies about $\omega+\omega_0$ and $\omega-\omega_0$, which will add together in a random way to produce a voltage of the form $V_n\cos\omega t\cos\omega_0 t$. It can be shown that the average value of the noise power transmitted in one direction by a wave guide is $kT\Delta f$, where Δf is the band width considered, and from this that the r.m.s. value of V_n is $(2kTZ\Delta f)^{\frac{1}{2}}$. This noise fluctuation produces fluctuations in the envelope of the power which are then comparable to those due to the gas absorption when the r.m.s. signal voltage $Vm/\sqrt{2}$ equals $(2kTZ\Delta f)^{\frac{1}{2}}$ or $(V^2m^2/2Z)=2kT\Delta f$.

The power transmitted through a wave guide absorption cell of length l is $P=P_0(\exp-(\alpha_c+\alpha_g)l)$ where α_g is the absorption coefficient of the gas. Therefore for minimum detectable signal, $(V^2m^2/2Z)=P_0e^{-\alpha_c l}(1-e^{-\alpha_g l/2})^2/4\approx2kT\Delta f$. Since we are interested in the case where $\alpha_g l\ll1$, $\alpha_{g\min}=$, $4/l[2kTe^{\alpha_c l}\Delta f/P_0]^{\frac{1}{2}}$. If α_g is minimized with respect to l, the optimum length $l_{op}=2/\alpha_c$, and then $\alpha_{g\min}=2e\alpha_c[2kT\Delta f/P_0]^{\frac{1}{2}}$. If a noise figure N is assumed for the detector or receiver, the apparent noise fluctuations are multiplied by the number N (by definition of N) and the expression for $\alpha_{g\min}$ becomes that given in Eq. (2). Although Eq. (2) has been developed for a particular type of spectrometer, it should be approximately correct for any instrument using a wave guide absorption cell, and readily extendable to other forms such as a resonant cavity spectrometer (also discussed by Hershberger).

As an illustration of the above relations, we take the following values for a microwave spectrometer for one-centimeter waves.

$\alpha_c=4\times10^{-4}$ nepers/cm or 8×10^{-4} cm^{-1},
$P_0=1$ milliwatt and
$\Delta f=30$ cycles.

The noise figure N is usually large and rather variable. We shall assume it to be 1 or zero db to illustrate the expected results under such ideal conditions. The optimum cell length is 25 meters, and the minimum detectable absorption coefficient at room temperature is 3.4×10^{-11} nepers/cm or 6.8×10^{-11} cm^{-1}.

Actual microwave spectrometers have not yet approached the theoretical performance allowed by fundamental noise fluctuations, but due to very large noise factors in the detectors and receivers and to various other interfering effects, they are at present generally limited to absorption coefficients larger than 10^{-7} cm^{-1}. For the far-from-ideal case where the predominating noise is from sources other than thermal fluctuations, detector, or receiver noise, or when the receiver noise figure varies rapidly with power, the above general discussion is of course no longer applicable and must be replaced by a consideration of the particular case.

* Work supported by the Signal Corps under Contract W36-039 sc-32003.

[1] W. D. Hershberger, J. App. Phys. **19**, 411 (1948).

Gyrograms

HENRY O. FUCHS
Preco Incorporated, Los Angeles 1, California
April 22, 1948

PROFESSOR ROJANSKY'S paper on "gyrograms" in the March issue of the Journal of Applied Physics draws attention to a simple and very useful method of integrating equations of the type:

$$m(d^2x/dt^2)+kx=f(t).$$

It may be interesting to note that the same method can be used without modification to integrate equations of the more general type:

$$m(d^2x/dt^2)+kx=f[t,\ x,\ (dx/dt)].$$

Professor Rojansky mentions the lack of references. An article on the subject was published in 1936.* In this reference examples were given showing the application of "gyrograms" to the solution of vibration problems with arbitrary damping, given as a function of velocity and displacement. The damping is, of course, equivalent to an external force.

The derivation shown by Professor Rojansky is practically the same as the one which I used to derive this method when I needed it about 1931 in an investigation of automobile shock absorbers. At that time, I was rather proud of it and showed it to my teacher who merely said, "Oh yes, an application of Proell's theorem!"

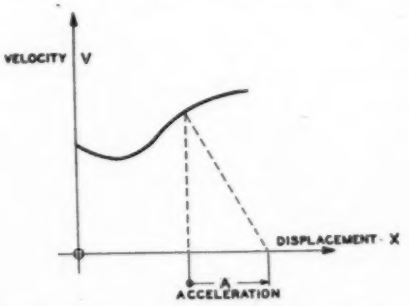


FIG. 1. Proell's theorem. Scales: Displacement: 1 inch represents a inches; Velocity: 1 inch represents b inch/sec.; Acceleration: 1 inch represents b^2/a inch/sec.^{2}.

A figure illustrating this almost unknown theorem is enclosed. The theorem states that in a displacement velocity diagram the subnormal is proportional to the acceleration and is easily proved by the relation:

$$\frac{d^2x}{dt^2}=\frac{dx}{dt}\frac{d}{dx}\left(\frac{dx}{dt}\right)=V\frac{dv}{dx}.$$

"Gyrograms" are superior to Proell's theorem because they specify a choice of scales which achieves two advantages: times are measured by angles, and the energy content of the system is measured by the distance from the origin.

* H. O. Fuchs, "Spiral diagrams to solve vibration damping problems," Product Engineering (1936).

A Note on Reflection from Dielectric Structures in Wave Guide*

J. Shmoys
Department of Physics, New York University, New York, New York
May 12, 1948

THE following method applies to dielectric structures of small cross section. This restriction can be more accurately stated as follows

$$(K_\epsilon-1)S\ll a\lambda, \tag{1}$$

where K_ϵ is the dielectric constant, S the area of obstacle, a the width of the guide, and λ the wave-length. The method consists of representing the effect of the obstacle in wave guide by the equivalent capacitive loading on a transmission line. The impedance of the transmission line is obtained readily from voltage in the center of the guide and power transmitted, and is given by

$$Z_0=(2b/a)(\lambda_g/\lambda)(\mu_0/\epsilon_0)^{\frac{1}{2}}, \tag{2}$$

where λ and λ_g are wave-lengths in free space and wave guide, respectively, b is the height of the guide, ϵ_0 and μ_0 are dielectric constant and permeability of free space. From this impedance the series impedance and shunt admittance per unit length of the equivalent line are obtained;

$$z=j(2\pi Z_0/\lambda_g) \tag{3}$$

and

$$y=j(2\pi/\lambda_g Z_0). \tag{4}$$

The loading capacitance is given by

$$dC=\epsilon_0(K_\epsilon-1)\sin^2(\pi x/a)(dS/b). \tag{5}$$

This is just the excess capacitance of a dielectric post in a parallel plate condenser if one takes into account the transformation ratio $\sin\pi x/a$ between the given point (x, z), and the center of the guide $(a/2, z)$, at which the voltage is defined. The validity of the method was suggested by the consideration of variational expressions obtained by Schwinger[1] for the impedance matrix elements of a dielectric post of arbitrary size and shape. The variational problem can be stated as follows:

$$\delta I(\phi)=\delta\left\{\frac{\int\phi^2(x,z)dS-2\pi^2/\lambda^2(K_\epsilon-1)\times\int\int\phi(x,z)G(x,z,x',z')\phi(x',z')dSdS'}{[\int\phi(x,z)\psi(x,z)dS]^2}\right\}, \tag{6}$$

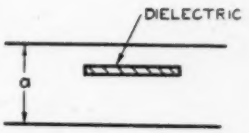


FIG. 1.

where ϕ is the unknown field in the dielectric, ψ the field in the undisturbed guide, G a Green's function and the integration extends over the obstacle. Two types of fields are considered, even and odd about $z=0$, yielding expressions for $(Z_{11}+Z_{12})$ and $(Z_{11}-Z_{12})$ of the four-terminal network.

The second term in the numerator of (6) is small compared to the first if (1) is satisfied, so that we have

$$I(\phi)=\{\int\phi^2dS/[\int\phi\psi dS]^2\}. \tag{7}$$

$I(\phi)$ minimizes then for $\phi=\psi$. This means that as a first approximation the electric field in the dielectric obstacle is that in the undisturbed guide. Hence the excess capacitance representation suggests itself. Schwinger showed that it is correct when the average dimension of the post is very small, which amounts to a confirmation of Eq. (5). The simplest application is, of course, to a thin rectangular block placed arbitrarily in the wave guide; for one position, when the block extends across the entire width of the guide perpendicular to the direction of propagation, the exact solution can be obtained. The approximate result proves to be the first-order term in thickness of the exact one.

It must be borne in mind that in neglecting the second term in Eq. (6) one automatically excludes from the analysis some discontinuity effects. Thus in the case shown on the diagram below, one simply gets a section of line of different impedance and propagation constant inserted into the wave-guide equivalent line. The inductive effects of the discontinuities at the back and front edges of the block, which could be shown to exist by a more detailed method, are here neglected entirely.

[1] Unpublished paper by J. Schwinger.
* This work was supported by the Air Materiel Command, Watson Laboratories, under Contract No. W28-099 ac-171.

The Theoretical Criterion for Streamer Advance in an Electrical Field

Leonard B. Loeb and Robert A. Wijsman
Department of Physics, University of California, Berkeley, California
April 2, 1948

THE streamer mechanism of spark breakdown in the higher pressure regions is an experimentally established fact.[1,2] Independently J. M. Meek[3] and H. Raether[4] tentatively proposed a semi-empirical condition for streamer advance which neglecting the important role of photo-ionization in the gas was fundamentally incomplete.[5] Recently L. B. Loeb[6] has indicated that the basic form of all sparking equations must bear a formal analogy to the original sparking equation of J. S. Townsend which is valid at low pressures, *viz.*,

$$\gamma e^{\alpha x}=1, \tag{1}$$

where γ is the coefficient involving the secondary process active and α represents the first Townsend, or primary ionization coefficient. This consideration led Loeb to the expression for the threshold of positive point, pre-onset burst, pulse corona[7] and in turn led to an indicated solution of the sparking threshold by streamer advance in a uniform field.[8] This approach to a solution was based on the assumption that streamer advance proceeds by the avalanche produced through *one photoelectron* advancing towards the initially created positive space charge in the sum of imposed and space charge produced electrical fields. While this solution may be legitimate in the case of uniform fields, it is certain that the pre-onset positive streamers from corona points in non-uniform fields[9] and perhaps streamers in uniform fields also can advance through the ionization by *several electron* avalanches feeding into the original positive space charge. This alternative solution is here outlined. In presenting this solution we shall use average values of quantities subject to considerable statistical fluctuation as is the case in all sparking theory. The solution cannot be put to quantitative test at present because of our ignorance of the values of the various quantities entering into the equation. In order to orient future consideration of the problem and indicate to experimental workers the data which are needed, the solution is here given.

To simplify presentation, the multiple avalanche mechanism will be analyzed assuming a plane parallel gap with a uniform field X. Adaptation to the non-uniform field will require little modification. In such a uniform field one electron leaving the cathode and advancing a distance δ towards the anode will create $e^{\alpha\delta}$ new electrons. These are absorbed into the anode leaving behind the same number of positive ions with greatest density at δ. These ions move slowly towards the cathode relative to the rapid motion of the electrons and will be considered at rest. At δ the ions lie within a radius ρ of the avalanche axis, given by

$$\rho=(4D\delta/v)^{\frac{1}{2}}, \qquad (2)^{10}$$

where v is the ionic drift velocity in the imposed field X, and D is the coefficient of diffusion. It has elsewhere been shown that the number N of ions formed in a volume $(4/3)\pi\rho^3$ of space charge at δ is given by

$$N=(4/3)\alpha\rho e^{\alpha\delta}. \qquad (3)^{10}$$

This leads to a space charge field X' at a distance r from the center M of the space charge of value

$$X'=(4/3)\epsilon\rho\alpha e^{\alpha\delta}/r^2=E/r^2, \qquad (4)$$

with ϵ the electronic charge and α the first Townsend coefficient. Thus the total field along the axis of the initial avalanche, on the side towards the cathode, will be

$$X_1=X+X'=X+E/r^2. \qquad (5)$$

The condition to be set for the extension of the positive space charge and its conducting plasma of transient electrons toward the cathode is that when N has reached an appropriate magnitude in the imposed field X of suitable magnitude, the photoelectrically initiated avalanches feeding into the positive space charge tip T at δ produce enough positive ions to maintain N and advance the tip.

Accompanying the ionization, yielding the N ions, there was produced a number f_1N photons. The factor f_1 may be taken as essentially constant. Some of these photons will liberate electrons in the surrounding gas by photoionization. Of these, some electrons will be created within a region R so close to the first avalanche axis, that their avalanches feed into T. It seems a fair approximation to take for this region R the space inside a coaxial cylinder, with cross-sectional radius ρ. The electric field at any point is $\mathbf{X}_1=\mathbf{X}+\mathbf{X}'$, the vector sum of $\mathbf{X}$ and $\mathbf{X}'$. For most of the points within R, the direction of $\mathbf{X}'$ is so close to $\mathbf{X}$, that X_1 can be taken as shown in Eq. (5).

If μ be the absorption coefficient of the photon, then in a region between r and $r+dr$, lying entirely in R, the number of absorbed photons is approximately $f_1N(\rho^2/4r^2)e^{-\mu r}\mu dr$. A fraction f_2 of these absorbed photons will lead to ionization, creating

$$f_2f_1N(\rho^2/4r^2)e^{-\mu r}\mu dr \qquad (6)$$

electrons between r and $r+dr$. Each of these electrons causes an avalanche of which the tip, at $r=\rho$, contains a number of ions equal to

$$(4/3)\bar{\alpha}'\rho'\exp\int_\rho^r \alpha' dr' \qquad (7)$$

in analogy to Eq. (3). Here α' is the first Townsend ionization coefficient in the field $X_1=X+[E/(r')^2]$. Use of r' in this integral instead of r is to indicate that this integration may be performed separately and preceeding later integration. The quantity $\bar{\alpha}'$ is the value of α' in X_1 when $r=\rho$. The symbol ρ' represents the radius of the new avalanche tip, which is smaller than ρ, as the electron travels a distance $r<\delta$. If the drift velocity is taken as approximately constant for small field changes, Eq. (2) fixes ρ' as

$$\rho'=\rho(r/\delta)^{\frac{1}{2}}. \qquad (8)$$

Combining (6), (7), and (8), and integrating from ρ to δ gives the total number of ions created by all avalanches near the tip T as

$$\int_\rho^\delta f_1f_2N(\rho^2/4r^2)e^{-\mu r}\mu dr\cdot 4/3\bar{\alpha}'\rho(r/\delta)^{\frac{1}{2}}\exp\int_\rho^r \alpha' dr'. \qquad (9)$$

This approximation is invalid if $1/\mu$ is not appreciably greater than ρ. The condition for streamer propagation is now clearly that the number of ions given by Eq. (9) be set equal to N, the number in the initial avalanche, so that on eliminating N the equation for the threshold of streamer propagation reads

$$\tfrac{1}{3}f_1f_2(\bar{\alpha}'\rho^3/(\delta)^{\frac{1}{2}}\mu)\int_\rho^\delta r^{-\frac{1}{2}}e^{-\mu r}\exp\int_\rho^r \alpha' dr' dr=1. \qquad (10)$$

This relation is seen to be of a type analogous to Eq. (1). It contains *all* the important factors though some are in averaged form. In contrast to Meek's equation, it will show through the variation of μ with p the greater departures from Paschen's law recently noted.[11] Solution of this relation for the sparking field strength X_s will follow the procedure pattern set by all previous theories beginning with that of Townsend's Eq. (1). Given are the stated values of p and δ and gas for the gap in question. Required

are the experimental values of the functions

$$\alpha/p = f(X/p), \tag{11}$$

for the gas together with values of the quantities f_1, f_2, μ, ϑ, and D which in function now replace the quantity γ of Eq. (1). In the small range of X/p covered in such a calculation, relation (11) can closely be approximated by a simple empirical analytical expression[12] permitting of integration in Eq. (7). A value of X is chosen by guess and through Eq. (2) ρ is evaluated. The value of α is then found from X, p and (11) and X' is determined from Eqs. (3) and (4). This yields the value of $\bar{\alpha}'$ and placed in the empirical form of (11) allows integration of the quantity (7). The evaluation of the left hand side of Eq. (10) is then possible. If this is different from unity, judicious new choices in X are made which permit the value of X_s for the solution of (10) to be accomplished by graphical means.

[1] H. Raether, Zeits. f. Physik **112**, 464 (1939).
[2] L. B. Loeb and J. M. Meek, J. App. Phys. **11**, 438 (1940).
[3] J. M. Meek, Phys. Rev. **57**, 722 (1940).
[4] H. Raether, Zeits. f. Physik **117**, 375, 524 (1941); Archiv. f. Elektrotechn. **34**, 49 (1940).
[5] L. B. Loeb and J. M. Meek, *The Mechanism of the Electric Spark* (Stanford Press, 1941), p. 41.
[6] L. B. Loeb, Rev. Mod. Phys. **20**, 151 (1948).
[7] L. B. Loeb, Phys. Rev. **73**, 798 (1948).
[8] L. B. Loeb, Phys. Rev. **73**, July 15, 1948.
[9] L. B. Loeb, J. App. Phys. (to be published).
[10] L. B. Loeb and J. M. Meek, *Mechanism of the Electric Spark* (Stanford Press, 1941), pp. 35, 37.
[11] C. G. Miller and L. B. Loeb, Phys. Rev. **73**, 84 (1948).
[12] L. B. Loeb, *Fundamental Processes Electrical Discharge in Gases* John Wiley, New York, New York, 1939), p. 347.

Notes on Electron Diffraction with the Universal Electron Microscope

R. E. Dunaway and H. M. Barton
Phillips Petroleum Company, Research Department, Bartlesville, Oklahoma

January 7, 1948

MODIFICATIONS have been described by Picard and Reisner[1] which allow the use of the RCA Model EMU electron microscope as an electron diffraction camera of effectively larger plate-to-specimen distances and higher resolving power than those attained using the procedure which is standard with this instrument. The sample is placed in the microscope stage, the objective and projector lens pole pieces are removed, and the usual condenser aperture is replaced by a smaller one so that the electron beam will have a sufficiently small angle of divergence at the specimen to produce patterns of high resolution. The pole pieces must be removed to prevent their obstructing the resultant image, as well as to decrease the magnetic fields of the lenses.

A similar method employing this instrument as an electron diffraction camera has since been successfully used in this laboratory. Instead of replacing the condenser aperture, the writers have placed another aperture on the microscope specimen holder by means of an adapter, as shown in Fig. 1. The aperture is provided for use with the regular diffraction holder which is normally inserted into the microscope column below the projector lens. This arrangement allows examination of the sample by microscopy, and after removing the pole pieces and intermediate viewing screen, by electron diffraction.

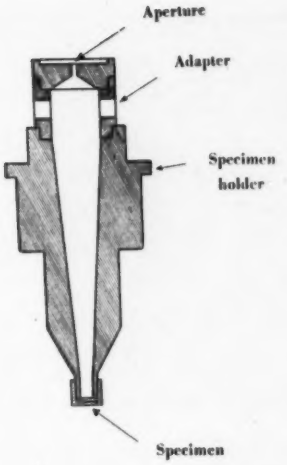


FIG. 1. Specimen holder with aperture and adapter.

The system must be realigned after removing the pole pieces in order that satisfactory diffraction patterns may be made, even though the electron-optical alignment may be correct for microscopy. Measurements show the ring diameters to be the same for patterns made under either condition of alignment but the line widths are larger for the misaligned conditions with a resultant decrease in resolution. The alignment procedure used for diffraction is that described by Picard and Reisner.[1] Some distortion may be introduced if the specimen holder does not receive the main electron beam, in which case an undesirable reflection of the beam above the specimen aperture serves as the source. Such patterns are characterized by a peculiarly shaped central beam spot as well as by non-uniformity of the ring widths. Measurements on patterns of evaporated gold films made at effectively short plate-to-specimen distances—i.e., low projector currents—indicated larger values for the product of ring diameters and corresponding lattice

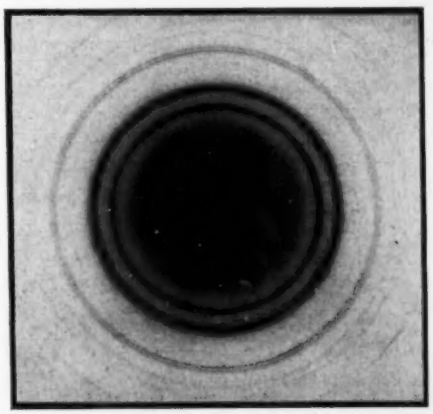


FIG. 2. Electron diffraction pattern of evaporated gold film with specimen below projector lens. Rings are elliptical.

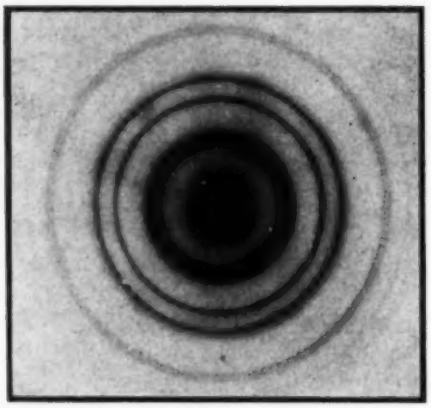


FIG. 3. Electron diffraction pattern of evaporated gold film with specimen in the microscope stage.

spacings for the larger diameter rings than for those near the center, whereas this product should be constant for any one diffraction pattern. At larger magnifications this product was found to be constant within the error of the measurements (about 1 percent). When transmission diffraction patterns are made using the regular diffraction holder placed below the projector lens the rings appear elliptical as shown in Fig. 2. With the specimen in the microscope stage the rings are more nearly circular as seen in Fig. 3, thus indicating little lens distortion in this respect.

The procedure for making magnified diffraction patterns, after having correctly aligned the electron-optical components, is to set the projector lens current until approximately the desired size pattern is seen on the final viewing screen. The objective lens current is then varied for the best focus with the condenser lens adjusted for the minimum size beam spot. In some cases it is necessary to move the specimen holder about to obtain more satisfactory patterns. Figure 4 shows ring diameters plotted against lattice spacings for different effective plate-to-specimen distances. The projector lens dial setting and the objective lens setting (dial readings for the FOCUS 1 selector switch) for various plate-to-specimen distances are also shown in Fig. 4. These calibrations have been reproduced at various

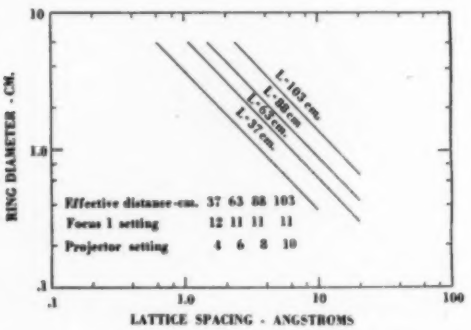


FIG. 4. Lattice spacings *vs.* ring diameters for various effective specimen-to-plate distances L.

times within 1 percent. With a 0.015-inch aperture in the specimen holder, the resolution obtained for an effective camera length of 63 cm is estimated to be 0.012A by measurements of the rings in gold patterns corresponding to spacings 0.935A and 0.910A, respectively.

The general method has the advantages of additional flexibility in ring sizes and the possibility of investigating larger spacings because of the reduction in background scattering near the central beam. The specimen aperture method (with a 0.015-inch aperture) gives less resolution than either the small condenser aperture or the standard method, but is advantageous to those who use the instrument primarily as a microscope and only occasionally for diffraction.

The specimen holder aperture method obviously suffers the disadvantage of limiting the motion possible in scanning a specimen. A better solution of the problem would be to provide some way to install the aperture somewhere between the condenser and the specimen holder in such fashion that it could be moved in and out of the beam as needed. Any portion of the specimen could then be examined by manipulation of the stage controls.

[1] Robert G. Picard and John H. Reisner, Rev. Sci. Inst. **17**, 484–489 (1946).

New Books

Vibration and Sound

Second Edition. By PHILIP M. MORSE. Pp. 468, McGraw-Hill Publishing Company, Inc., New York-Toronto-London 1948. Price $5.50.

Here is a masterful revision of a book that already has won its place as an outstanding contribution to the literature of acoustics. The author is well known for his contributions to theoretical work not only in acoustics but also in many other branches of physics, and an outstanding feature of the present volume is the application of mathematical methods hitherto not generally used on acoustical problems. The first edition of this work applied to the problems of wave acoustics the methods that had proved so powerful in the Schrödinger form of the quantum theory. In addition to extending this treatment in the present edition, we find a more complete development of first-order perturbation methods illustrated in problems involving the non-linear string and non-linear bar, and more particularly in room acoustics where they are used to give the approximate behavior of non-rectangular rooms and rooms having irregular structural features.

The outstanding addition to the new edition is a thorough revision of practically every chapter to include in it a treatment of transient phenomena. As ground work the author has included in Chapter One a delightfully simple introduction to the subject of contour integration as the basis for the subsequent evaluation of Fourier and Laplace transforms that arise in the treatment of transients.

Especially noteworthy among the types of transients considered are those encountered in the general analysis of the motion of the vibrating string, the transients in the motion of the thin diaphragm, when used in the condenser microphone, a beautiful treatment of the troublesome flutter echo, and a clear exposition of the place that transients occupy in the performance of horns, and the transmission of sound through ducts. The section on the propagation of sound through ducts is developed by methods standard for electrical network and transmission line analysis.

The new edition has added the accepted method of acoustical analogies for certain important applications, specifically, in dealing with the basic problem of the acoustical properties of constrictions and holes needed for the analysis of the behavior of sound absorbing materials, in the behavior of various wall structures, and in describing the action of horns. In the latter section the pioneer work of Salmon on catenoidal horns is included as an important addition.

The chapter on the radiation and scattering of sound is as outstanding in this edition as it was in the first edition. This chapter is a definitive exposition of present knowledge and methods in this field with the exception of the powerful variational methods recently introduced into acoustical problems by Schwinger.

Perhaps the most outstanding contribution of this work is the chapter dealing with room acoustics. Throughout the entire treatment of this subject, the development emphasizes the basic importance of the acoustic impedance of the wall materials, although the more familiar treatment in terms of absorption coefficients is also developed—and so skillfully that workers long familiar with this method will readily appreciate its relationship to more basic approach by the acoustic impedance method. The absorption and reflection of sound at a surface are treated in a thorough-going and complete manner from this viewpoint in terms of the acoustic impedance of the boundaries. The normal modes of vibration of the air in enclosures is considered both from the approximate solution and by the more exact methods used by Morse and Bolt and others. The behavior of such enclosures under steady-state forced vibration is thoroughly presented, and this chapter presents a most complete description of the reverberation properties of rooms in terms of transient response and decay after either forced vibration or impulse excitation. In the treatment of the latter aspect of this problem, the methods of transient analysis featured in this volume find some of their most important applications.

A review of this book would not be complete did it not pay tribute to the fine tables and graphs of functions which the author has prepared for the readers' convenience. The volume also includes an exceptionally fine group of problems which not only illustrate the text discussions, but indicate methods of solution for some of the principal problems in acoustics. There is no question that this book will long occupy a central position for workers in acoustics.

ROBERT S. SHANKLAND
Case Institute of Technology

New Booklets

Westinghouse Electric Corporation, Technical Press Service, 306 Fourth Avenue, Pittsburgh 30, Pennsylvania, has announced the availability of a *Table of Interplanar Spacings in Angstrom or K.X. Units in Terms of 2θ for Different Target Materials*. 18 pages, 1 figure. Free on request.

James G. Biddle Company, 1316 Arch Street, Philadelphia 7, Pennsylvania, publishes *Biddle Instrument News*, a four-page folder of news items from the electrical industry. Free on request.

Century Geophysical Corporation, 1333 North Utica, Tulsa, Oklahoma, has published two bulletins on its portable seismograph equipment and galvanometers: *Century Portable Recording Equipment* and *Century Galvanometers*. 8 and 4 pages, respectively. Free on request.

The Baldwin Locomotive Works, Testing Equipment Department, Philadelphia 42, Pennsylvania, offers a free 4-page bulletin (No. 282) describing and illustrating Baldwin-Tate-Emery universal testing machine models.

Eastman Kodak Company, Sales Service Division, Rochester 4, New York, announces two new pamphlets: one, 16 pages, lists references on types of photomicrography and their uses; the other, 32 pages, lists references on fields of application of both photomicrography and electron micrography. Free on request.

The Harshaw Chemical Company, 1945 East 97th Street, Cleveland 6, Ohio, has published a 32-page booklet entitled *Industrial Chemicals*. It lists all chemicals available through the company, briefly relates the history of the organization, and describes its major activities. Free on request.

Here and There

New Appointments

Simon Ramo, formerly of General Electric Company, has been appointed director of guided missile development by the Hughes Aircraft Company, where he has been acting in the capacity of director of research for the electronics department. He has also recently been appointed a research associate at the California Institute of Technology, where he teaches a doctorate course in advanced radio engineering theory.

W. W. Mutch will be on leave of absence from Knox College next year to serve as acting head of the physics department at Wabash College. He replaces Duane Roller,

who will be at Harvard University on a visiting appointment. Clayton M. Zieman has been appointed assistant professor of physics at Wabash, effective in September.

George R. Bastedo has been appointed to the staff of the National Bureau of Standards, where he will work as electronics engineer for the Guided Missile Section.

Harold Vagtborg has resigned as president and director of Midwest Research Institute to become president and director of Southwest Research Institute, effective September 1. He will also serve as director of three other research organizations associated with the Houston, Texas, institute.

Awards

The following physicists are among a group of 65 scientists and engineers of the wartime Office of Scientific Research and Development who received the Medal of Merit recently in recognition of their services to the armed forces: Luis W. Alvarez, Lee A. DuBridge, Lars O. Grondahl, Zay Jeffries, Paul E. Klopsteg, Irving Langmuir, Carroll L. Wilson.

Two staff members of the National Bureau of Standards have been presented awards. Robert D. Huntoon, assistant chief of the atomic physics division, received one of the two distinguished achievement awards given by the Washington Academy of Sciences for 1947. He was cited for his research in "the advancement of electronics and its application to other sciences and to modern ordnance." A. V. Astin, assistant chief of the electronics division, was awarded His Majesty's Medal for Service in the Cause of Freedom in recognition of "valuable services rendered to the Allied War effort." Dr. Astin was in England from September 1944 to March 1945 as representative of the National Bureau of Standards and of the National Defense Research Council. As Government officials are not permitted to accept foreign decorations, the medal will be held in custody by the Department of State until Dr. Astin's retirement from Government service.

Northwest Conference on Nuclear Science

Reed College, Portland, Oregon, held its first Northwest Conference on Nuclear Science June 28 to July 16. It was planned to provide technical study for college and research laboratory scientists, a survey for high school science teachers, and orientation for laymen. The faculty included three staff members of the University of Chicago's Institute for Nuclear Studies—Samuel K. Allison, Anthony Turkevich, and John A. Simpson, Jr.

Nuclear Physics Course at UCLA

From August 2 through August 20 a course in the application of nuclear physics to the biological and medical sciences is being given at the University of California at Los Angeles. It is under the direction of Stafford L. Warren, M.D., and Fred A. Bryan, M.D., of the School of Medicine of the University.

Atomic Energy Project at Union College

Study of the biological effects of high-energy radiations, such as 100-million-volt x-rays, will be undertaken by Union College in a research project financed by the Atomic Energy Commission for the Office of Naval Research. The scientific work, which will be directed by Dr. Leonard B. Clark, chairman of the college's Department of Biology, is expected to start this summer. The General Electric Company will make available at its own expense certain resources of its research laboratory, including use of the 135-ton betatron, which will be used to radiate the biological specimens.

EMSA Placement Service

The Electron Microscope Society of America has set up a placement service for the mutual benefit of its members and potential employers. The Society has on its membership roll the names of virtually every active electron microscopist in this country. It is suggested that employers seeking scientists or technicians for work in this or allied fields make use of the placement service. Inquiries in this regard should be directed to the Secretary of the EMSA, Dr. C. J. Burton, American Cyanamid Company, 1937 West Main Street, Stamford, Connecticut.